Maria Laura Bettinsoli and Caterina Suitner

Wandering in Social Cognition: a Journey into Shared Realities and Intergroup Dynamics

Wandering in Social Cognition: a Journey into Shared Realities and Intergroup Dynamics

edited by

Maria Laura Bettinsoli and Caterina Suitner

Table on Contents

Introduction

The authors of this book are Anne Maass' friends, former pupils, and close colleagues, writing this book from all over the world with a twofold goal: celebrating our connection with Anne and establishing a common approach to social cognition, inspired by Anne's work. She affects not only our questions and methods, but also our heart: her passion for collaborative science spreads virally to any scholar who is lucky enough to collaborate with her. Anne retired from the University of Padova in 2022 to start a brand-new career at NYU Abu Dhabi, where she is currently happily working. This significant passage gave us the perfect excuse to explicitly state and honor the central role Anne represents in our scientific adventure in the social cognition landscape. Gratitude is the common emotion linking all of us to Anne, and to each other.

The central focus of this book lies in exploring two fundamental social processes in social cognition. First, it delves into the construction of a shared reality; second, it examines the formation of categories and their consequences on intergroup dynamics. Both processes are highly rooted in the fundamental activity of encoding (and decoding) our social world in and beyond space and time through language, the fil rouge sewing this book, and their authors to Anne Maass.

The theoretical chapter authored by Kashima (Chapter 1) serves as a foundational introduction to the process behind the creation of social knowledge, elucidating how social cognition is deeply rooted in the creation of a shared culture. Specifically, the chapter illustrates how this shared reality is shaped through the sophisticated use of language devices. In fact, language - as the primary tool of human communication - enables individuals to convey and interpret a collective story. Further elaborating on the central role of storytelling for establishing a common understanding of social reality, Castano (Chapter 2) emphasizes the pivotal role of literary fiction in building our ability to grasp others' feelings, attitudes, and beliefs (i.e., Theory of Mind). In fact, through storytelling, humans train their ability to understand the mental states of others, by adopting different perspectives. Established that this ability is honed through a linguistic code (literary fiction), a further question pertains to whether the languages we speak matter in the emergence of our emotional intelligence. In other words, are we always able to read others' minds with the same efficacy across different linguistic codes? Foroni (Chapter

3) sheds light on how our emotional competence is honed through language, with our native language (L1) being more effective than other languages (L2s) in dealing with ours and others' emotions. In summary, language serves as a vital tool for encoding the cultural building blocks that shape our shared reality, whether specific to a particular culture or universally applicable. This interplay between culture and social cognition is specifically addressed by Goyal (Chapter 4), who highlights how culture can indeed introduce cognitive biases, particularly in memory, as specific content might become more salient in a particular language/culture than in others, thus shaping our recollection. Karasawa (Chapter 5) further explores this concept, emphasizing that agentivity—the expression of agency—is more robustly conveyed in certain cultures and languages. Essentially, language serves as the vessel for encoding critical cultural elements. Agency is indeed a pivotal dimension as it signals the ability to act independently, achieve goals, and make choices. Durante (Chapter 6) delves into agency by describing a key model in the realm of social perception — that is, the Stereotype Content Model, which provides a comprehensive explanation of how we perceive and comprehend social dynamics. But how does agency manifest in language? As explained by Formanowicz & Suitner in Chapter 7, the Social Grammar Model clarifies that language uses different mechanisms to represent agency, for instance with the use of verbs. However, encoding and decoding agency is not limited to grammatical elements, but it can be also related to specific syntactic structures. For instance, the order in which we arrange words in a sentence (Bettinsoli, Chapter 8) might drive the attribution of responsibility of an action but also how we infer causal relations. However, these processes go also beyond verbal language to enter the realm of visual and spatial representation. One of the social cognition topics mostly investigated by Anne Maass is the Spatial Agency Bias, which demonstrates how word ordering translates into visual representation. Written text symbolically captures agency within the visual space (Suitner & Formanowicz, Chapter 9). Anne Maass has extensively studied the Spatial Agency Bias, a social cognition phenomenon that shows how the order of words in language affects how we perceive them visually. The direction we write in and where we place elements in a text reflect how we assign agency in the visual space (see Suitner & Formanowicz, Chapter 9). For instance, groups associated with agency tend to be symbolically placed on the left side of a picture, a text, or a map, because the left side is often perceived as more prominent, dominant, or active than the right

side for left-to-right language speakers. Let's consider gender: women, often stereotypically linked to lower agency, are positioned to the right in graphs and diagrams as discussed by Hegarty in Chapter 10.

An intriguing aspect related to the interplay between the perception of psychological distance and agency is closely tied to our self—the most immediate form of identity; and this process is well explained by Liberman and colleagues (Chapter 11) who guide us into the caveats of the Construal Level theory by exploring its potential commonalities and divergence with Spatial Agency Bias. In summary, language acts as a multifaceted tool, encoding cultural nuances and shaping our perception of agency across various contexts. As we navigate into these chapters, we will realize that this consistent pattern of encoding agency through various tools remains culturally specific.

The second process under which we can conceptually cluster the remaining chapters of this book refers to the formation of categories and their consequences on intergroup dynamics. One significant challenge in social psychology is comprehending the cognitive pathways leading to a highly unequal social hierarchy where power and resources are controlled by high status groups at the expense of the others, with dire consequences, well elucidated by Cervone and colleagues (Chapter 12). Anne Maass has consistently focused on advocating for social minority groups since the beginning of her research career (as discussed in Volpato & Mucchi Faina's Tribute). A central question explored by many scholars, alongside Anne Maass, pertains to understanding the cognitive processes that contribute to the establishment of intergroup discrimination, derogation and social conflict. Chapter 13, as framed by Hamilton, delves into the fundamental mechanisms through which we construct descriptions of social categories, namely the building block of social relations. By observing behaviors and instances in our social environment, we infer generalizable traits from concrete examples through an abstraction process that ultimately defines stereotype content. Hamilton elucidates the fundamental principles underlying the creation of stereotype content. This process is rooted in one of our innate cognitive biases—that is, an upward tendency that shapes our perception of reality. Essentially, we spontaneously start from specific examples and instances to ultimately summarize them into broader and more abstract categories. These categories further define the traits (and stereotypes) attributed to individuals and groups, in a circular way. Hamilton's work, along with insights from Castelli & Carraro (Chapter 14), highlights how we are socialized through exposure to this

content and how we learn since early age to strategically use this content to favor our ingroup while discriminating against outgroups—basically, categorization becomes a tool for derogation. Salvador Casara (Chapter 15) further explores how basic cognitive biases are enriched by social motives, contributing to shape our intergroup relations, and how intergroup drives often overrule the rationality of our evaluations. Along this line is the "gaydar" phenomenon (Fasoli, Chapter 16), a mainly inaccurate categorization of speakers' sexual orientation that leads to derogation due to underlying motivational elements. The central focus of these chapters stays on the crucial role that social motives play in shaping our cognitive architecture, an interconnected net of attitudes, beliefs, and cognitive functions. A net that is effectively exploited by implicit measures of social cognition to capture social processes, as elucidated by Marini (Chapter 17). As such, even vocal or facial features can be investigated to detect discrimination. For instance, Dalmaso and colleagues (Chapter 18) showed that individuals tend to be less influenced by faces belonging to ethnic groups associated with a lower social status. Similarly, the stereotype of women as having lower status or capabilities (for example in navigating space, Pazzaglia et al., Chapter 19) hamper their progress in contexts like academia (Galdi, Chapter 20), defining the inequality that characterizes our society when inspected through a gender lens. A key process in the maintenance of social hierarchies is to treat minority members as less than human. The deprivation of humanity is critical to sustain and justify social inequality, as it reduces empathy and solidarity. Three instances of this dehumanization are outlined here. Carnaghi and Bianchi (Chapter 21) illustrate different theoretical perspectives on how derogatory group labels contribute to the denial of humanity for social minority groups. Andrighetto & Baldissarri (Chapter 22) focus on the phenomenon and consequences of dehumanizing working experiences, showing the costs of objectification for workers' well-being. The third feature of dehumanization is addressed by Guizzo (Chapter 23), focusing on the detrimental effects of sexual objectification on women's self-perception and well-being, among which of the most dangerous and extreme consequences of sexual objectification, namely, sexual harassment and violence.

Throughout this book and in Anne Maass' approach, linguistic cues are investigated to understand socio-cognitive processes (e.g., Kashima, Chapter 1; Suitner & Formanowicz, Chapter 9), but they can also be exploited as carriers of social change. In Chapter 24, Merkel and Roessel shed light on the pivotal role of language in shaping gender norms.

Bruckmüller (Chapter 25) delves into the significance of framing economic inequality, potentially dismantling established social orders.

This book represents a collective effort to contribute with gratitude to the scientific challenges that Anne Maass has opened within the realm of social psychology. These challenges stimulated the authors in diverse and idiosyncratic ways, with the common outcome that each of us feels that Anne Maass has been shaping the style, the content, the trajectories of our own scientific endeavors, and our discipline.

We are all eternally grateful to Anne for her immeasurable contributions, both to the field and to each of our personal journeys as scholars. In the spirit of that gratitude, we hope to give back, if only a little, by surprising her with this book—an offering that makes visible the vast array of research and scholarship she has inspired.

Anne Maass and the investigation of minority influence: when and where everything started

Chiara Volpato[1] and Angelica Mucchi Faina[2]
[1]University of Milano-Bicocca, Italy
[2]University of Perugia, Italy

Minority influence was Anne Maass's first scientific love. When Anne went to the United States to pursue her doctoral studies at Florida State University, she brought with her the idea of studying minority influence. She wished to delve into theoretical and research ideas proposed in France in the late 1960s by Serge Moscovici (1976, 1980) and unknown in the United States, which was still tied to studies of conformism and majority influence. In Tallahassee, Anne completed her doctoral studies under Russell Clark, with whom she devoted herself to understanding the processes and functioning of active minorities.

Maass and Clark's early studies on minority influence were published in the European Journal of Social Psychology, a journal established in 1970 at the behest of leading European social psychologists who wished to create a reflection hub characterized by greater "social" engagement of the discipline.

Her first research (Maass, Clark, & Haberkorn, 1982) investigated if Moscovici's theory applied to the social minorities in the United States at the time of the study. She and her colleagues examined how either perceived competency or self-interest and Zeitgeist affected minority influence. The study focused on two minorities: The pro-abortion group (i.e., pro-Zeitgeist) and the death penalty group (i.e., anti-Zeitgeist). The self-interest notion predicts that "single" minorities (deviating only in terms

of beliefs) are more influential than "double" minorities (also deviating in category membership) and, further, that either minority is influential only when the Zeitgeist is in favor of the minority position. The results demonstrated that "double" minorities were perceived as having a stronger self-interest and exerted less influence than single minorities and that the Zeitgeist played a crucial role in making a minority influential.

In the following research (Maass & Clark, 1983), two experiments investigated whether minority influence and conformity operated by the same or different processes. It was predicted that subjects who were simultaneously exposed to a majority and a minority opinion would move towards the minority in private but towards the majority in public. The results supported the hypothesis, confirming the dual process model.

In 1984, Maass and Clark published work in the Psychological Bulletin, which has become an essential reference in this area of research. After highlighting several theoretical (i.e., scarcity of theoretical integrations, ignorance of the psychological processes underlying the Zeitgeist phenomenon) and methodological (i.e., inconsistent operationalizations of key concepts, the omission of process-oriented methodologies, inadequate treatment of mediating variables and group effects) problems that could be found in the previous studies, they affirmed the importance of applying the theory of minority influence to actual minority groups. Also, they suggested that the relationship between minority influence and conformity needed to be further explored.

In the following years, Maass and Clark continued to work together in this area of research. In 1986, the authors performed an experiment to test whether reactance theory could account for private acceptance of a minority opinion under simultaneous majority/minority influence (reactance against majority rather than conversion towards minority), ruling out reactance as an alternative explanation to the theory of minority influence proposed by Moscovici (1976, 1980).

In 1988, Clark and Maass published two articles concerning the role of social categorization in minority influence. In the first paper (Clark & Maass, 1988a), they presented, in the framework of Tajfel's social identity theory, three experiments in order to compare the influence of ingroup and outgroup minorities and to assess the role of Zeitgeist perception in minority influence. They confirmed once again that ingroup minorities were more influential than outgroup minorities. This finding was observed in two different experimental paradigms, using either a small group setting in which subjects interacted with the minority or the si-

multaneous social influence paradigm in which both influence sources impinged simultaneously upon the subjects.

The second work (Clark & Maass, 1988b) was conducted in order to compare the influence of ingroup and outgroup minorities and to assess the role of perceived source credibility in minority influence. Participants were exposed to the simultaneous majority/minority influence paradigm. They moved towards the minority position in private and the majority position in public when members of the ingroup represented the minority. In private responses, participants were not affected by outgroup minorities who argued for abortion, and they became more positive towards abortion when outgroup minorities opposed abortion. Ingroup minorities were perceived as more credible than outgroup minorities, and greater credibility of minority sources was associated with greater attitude change towards the minority position.

Also, Clark and Maass (1990) conducted two experiments in order to investigate the role of majority size (social pressure) in minority influence. In Experiment 1, there was a tendency for minority influence to decrease with increasing majority size when the minority argued against gay rights. The results were stronger in Experiment 2. For pro-abortion minorities, the minority's impact declined as the size of the opposing majority increased. As predicted by the social impact models, this decline occurred during the initial increase of the majority size.

Anne Maass conducted the last studies on minority influence in the United States with Melanie Trost and Douglas Kenrick (Trost, Maass, & Kenrick, 1992; see also Trost & Kenrick, 1994). They examined the role of personal relevance, which can set an important boundary condition on minority influence, limiting the persuasiveness of consistent minority advocacy on personally essential issues. Previous research has found that a minority group's advocacy elicits private acceptance of its message. However, these findings derived from studies in which minorities advocated issues of low to moderate personal involvement for subjects. Theoretical considerations from the persuasion literature led to the hypothesis that high personal relevance could elicit rejection of a minority influence attempt. Maass and colleagues' results showed that a minority source was persuasive when advocating a counter-attitudinal position of low personal relevance. Under conditions of high relevance, however, the minority's arguments elicited only resistance. Positive responses to minorities under conditions of low relevance were replaced with minority

derogation under high relevance. The majority advocacy elicited undifferentiated responses, regardless of the level of personal relevance.

In later years, Anne Maass met with European researchers working on active minorities. The first approach occurred during a workshop held in Geneva in 1985 at the department of Willem Doise and Gabriel Mugny, followed by a conference in Perugia in 1989, organized by Angelica Mucchi Faina, and a meeting in Valencia in 1990. These collaborations resulted in the book Minority influence, edited in 1994 by Serge Moscovici, Angelica Mucchi Faina and Anne Maass. The volume was the first collection of minority influence research in which there was a balance between the contributions of authors from Europe (Fabrizio Butera, Claude Kaiser, Anne Maass, Serge Moscovici, Angelica Mucchi Faina, Gabriel Mugny, Juan Pérez, Bernard Personnaz, Marie Personnaz, Giovanna Petrillo, Patricia Roux, Chiara Volpato, Erich Witte) and the United States (Russell Clark, Dawna Coutant, William Crano, Michele Grossman, Douglas Kenrick, Charlan Nemeth, Melanie Trost, Stephen Worchel).

The book also testifies to the harmony that, towards the end of the 1980s, had developed between Anne and the authors of the present contribution. We had begun to hang out and do research together on the topic of minority influence, which had fascinated us not only because of its absolute novelty – given its contrast with the classic paradigm of conformism – but also because of its possible political and social impact. We taught and lived in different cities, so we met in Bologna and spent the day in the comfortable lobby of a hotel opposite the station, where they allowed us to work and served us delicious sandwiches. There, we devised the experiments that we would publish in later years.

Our ideas were built from the comparison between the conversion theory of Moscovici (1976, 1980, 1994) and the divergence theory of Nemeth (1986, 1994). These two models have received considerable empirical support suggesting that both approaches make valid predictions under specific circumstances. According to Moscovici's conversion theory, the majority and minority operate different processes of influence: The majority operates a comparison process in which the various opinions are compared without further attention to the stimulus under discussion. In contrast, the minority operates a validation process, which needs the examination of the relation between the minority's response and the object or reality. Majority influence is likely to manifest in public compliance, which is limited to the influence setting and persists only as long as the influence source remains present. In contrast, the deeper

processing of the minority message produces a private and enduring conversion that will persist in the absence of the minority, although it may not surface in public.

According to the model of Nemeth (1986), the majorities induce convergent thought processes in which people focus their attention on the majority's message without considering further aspects of the issue under discussion, while the minority groups induce divergent thinking, focusing on a wide stimulus array well beyond the minority's message. Consequently, people tend to follow the majority regardless of whether the majority is correct or incorrect. In contrast, they are unlikely to adopt the minority position. Still, they tend to detect new, alternative solutions that had not been proposed by the minority and that, in the absence of minority influence, would have gone undetected. Since these novel alternatives tend to be correct, minorities stimulate creativity and improve performance (Maass & Volpato, 1991, 1994).

In our opinion, the theories of Moscovici and Nemeth were not alternative but complementary. For this reason, it was essential to understand which situational and personality constellations were likely to produce conversion and which were likely to produce divergence effects. This research question was challenging both empirically and theoretically. From a theoretical perspective, it was important to understand when minorities would create movements towards their position (conversion) and when they would stimulate divergent thought leading to new, original solutions. From a methodological perspective, a test of the Nemeth theory would require the development of new methodological tools. We advanced some tentative hypotheses about when divergence or conversion effects would occur. In particular, we argued that divergence effects would most likely emerge when personal relevance was high, in a setting that fostered creativity, and when people had a high probability of engaging in and enjoying effortful cognitive endeavors.

Our first work (Volpato, Maass, Mucchi Faina, & Vitti, 1990) regarded the social categorization in the minority influence and proposed a theoretical distinction between representative outgroup minorities (representative of a minority category in the society, e.g., gays) and dissident outgroup minorities (defined as a minority subgroup within a larger outgroup category). The article described two studies comparing the social influence of dissident outgroup minorities with that of ingroup minorities (belonging to the participant's own social category). The conflict between minority and majority within one's own group should be particu-

larly relevant to the participant, whereas a similar disagreement within the outgroup is personally less involving. We, therefore, predicted that the divergence effects would occur when people were exposed to influence from an ingroup minority. In contrast, conversion effects were more likely to emerge when confronted with a dissident outgroup minority. We exposed high-school students from Milan to a minority advocating a revision of the final high-school examination intended to further increase the examination standards. The minority was either described as coming from the participant's city (i.e., Milan) or Rome (two traditionally antagonistic cities, with Milan representing economic power and Rome political power). Participants were informed that the Ministry of Education had asked two student committees, one from Milan and one from Rome, to evaluate the proposed revision of the examination. They either received the alleged minority report from the Milan committee (i.e., ingroup) or the Rome committee (i.e., outgroup). In either case, participants believed that the committee had presented two separate reports, a majority and a minority report and that they would receive only the minority report, which was in favor of the new examination. After reading the minority report, participants were asked to privately indicate their own opinion by selecting one of three response options: Agreement with the minority (conversion measure), agreement with the current regulation, or the proposal of a new alternative (divergence measure). The results strongly supported our predictions. People tended to adopt the position advocated by a dissident outgroup minority, but rather than going along with an ingroup minority, they generated new alternative solutions. Dissident outgroup minorities induced a conversion effect, whereas ingroup minorities stimulated the divergent thought processes outlined by Nemeth's theory. Interestingly, this pattern dramatically changed when people were confronted with a majority source. In a second experiment, subjects were exposed to a majority or to a minority source, either belonging to the subject's social category or to the outgroup. The results indicate that the position of an ingroup majority was readily accepted, whereas the otherwise identical message of an outgroup majority was rejected; neither the ingroup nor outgroup majority stimulated the development of alternative proposals. Again, in line with Nemeth' (1986) theory, the position of an ingroup minority was rejected but stimulated the generation of new, alternative proposals.

In our second work (Mucchi Faina, Maass, & Volpato, 1991), we started from the idea that a minority advocating an original position may

introduce a new rule, replacing correctness with originality. An experiment illustrated how an original minority, unlike a conventional minority, might stimulate divergence. Participants were asked to provide suggestions on "how to promote the image of Perugia internationally." They were simultaneously exposed to a majority and a minority source of influence. The majority proposed a conventional representation of Perugia (a photograph of Palazzo dei Priori). The minority was either described as having made an alternative but equally conventional proposal (a photograph of Arco Etrusco) or an original proposal (a photograph of the escalator underneath the Rocca Paolina, representing a modern element in a medieval city). Besides privately expressing their agreement with the majority and minority proposals, participants were also asked to indicate additional proposals that came to their minds. The results were clear: the original minority stimulated greater originality. It is important to note that neither originality per se nor minority influence per se stimulates divergent thought processes, but rather the unique combination of the two. A second experiment further demonstrated that the original message induced creative processing only when attributed to a minority source but not to a majority source. Taken together, the results suggested that the divergence effects were particularly likely to emerge in settings in which originality was stressed.

The third article (Maass, Volpato, & Mucchi Faina, 1996) presented two experiments comparing majority and minority influence on attitudinal (opinion) vs. objective (knowledge) tasks. The hypothesis that minority influence would decline on objective items was tested by exposing participants to a minority or majority influence source. The question under discussion was either objective ("from which country does Italy import most of its raw oil?") or attitudinal ("from which country should Italy import most of its raw oil?"). The first experiment showed that, compared to a non-influence control group, majorities strongly influenced both objective and attitudinal issues. In contrast, minorities were persuasive only on attitudinal issues. The second experiment indicated that this was true only for participants who were uncertain of their own position, while minorities were unable to convince highly certain participants regardless of the type of task.

References

Clark, R. D., & Maass, A. (1988a). Social categorization in minority influence: The case of homosexuality. *European Journal of Social Psychology*, 18(4), 347-364.

Clark, R. D., & Maass, A. (1988b). The role of social categorization and perceived source credibility in minority influence. *European Journal of Social Psychology*, 18(5), 381-394.

Clark, R. D., & Maass, A. (1990). The effects of majority size on minority influence. *European Journal of Social Psychology*, 20(2), 99-117.

Maass, A., & Clark, R. D. (1983). Internalization versus compliance: Differential processes underlying minority influence and conformity. *European Journal of Social Psychology*, 13(3), 197-215.

Maass, A., & Clark, R. D. (1984). Hidden impact of minorities: Fifteen years of minority influence research. *Psychological Bulletin*, 95(3), 428-450.

Maass, A., & Clark, R. D. (1986). Conversion theory and simultaneous majority/minority influence: Can reactance offer an alternative explanation? *European Journal of Social Psychology*, 16(3), 305-309.

Maass, A., Clark, R. D., & Haberkorn, (1982). The effects of differential ascribed category membership and norms on minority influence. *European Journal of Social Psychology*, 12(1), 89-104.

Maass, A, & Volpato, C. (1991). Prospettive teoriche sull'influenza minoritaria: Conversione o divergenza? Ricerche di Psicologia, 4, 9-23.

Maass, A, & Volpato, C. (1994). Theoretical perspectives on minority influence: Conversion versus divergence? In S. Moscovici, A. Mucchi Faina, & A. Maass, A. (Eds) (1994). Minority influence (pp. 135-147). Chicago: Nelson-Hall.

Maass, A, Volpato, C., & Mucchi Faina, A. (1996). Social influence and the verifiability of the issue under discussion: Attitudinal vs objective items. *British Journal of Social Psychology*, 35, 15-26.

Moscovici, S. (1976). Social influence and social change. London: Academic Press.

Moscovici, S. (1980). Toward a theory of conversion behaviour. In L. Berkowitz (Ed.) Advances in Experimental Social Psychology, vol.13, pp. 209-239. New York: Academic Press.

Moscovici, S. (1994). Three concepts: Minority, conflict, and behavioural style. In S. Moscovici, A. Mucchi Faina, & A. Maass, A. (Eds) (1994). Minority influence (pp. 233-251). Chicago: Nelson-Hall.

Moscovici, S., Mucchi Faina, A., & Maass, A. (Eds) (1994). Minority

influence. Chicago: Nelson-Hall.

Mucchi Faina, A., Maass, A., & Volpato, C. (1991). Social influence: The role of originality. *European Journal of Social Psychology*, 21, 183-197.

Nemeth, C. J. (1986). Differential contributions of majority and minority influence. Psychological Review, 93, 23-32.

Nemeth, C. J. (1994). The value of minority dissent. In S. Moscovici, A. Mucchi Faina, & A. Maass, A. (Eds) (1994). Minority influence (pp. 3-15). Chicago: Nelson-Hall.

Trost, M. R., & Kenrick, D. T. (1994). Ego involvement in the minority influence paradigm: The double-edged sword of minority advocacy. In S. Moscovici, A. Mucchi Faina, & A. Maass, A. (Eds) (1994). Minority influence (pp. 149-161). Chicago: Nelson-Hall.

Trost, M. R., Maass, A., & Kenrick, D. T. (1992). Minority influence: Personal relevance biases cognitive processes and reverses private acceptance. *Journal of Experimental Social Psychology*, 28(3), 234-254.

Volpato, C., Maass, A., Mucchi Faina, A., & Vitti, E. (1990). Minority influence and social categorization. *European Journal of Social Psychology*, 20, 119-132.

1. Cognition, Communication, and Social Reality

Yoshihisa Kashima
Melbourne School of Psychological Sciences, The University of Melbourne, Australia

Imagine a conversation between aliens on the planet Twin Earth. Suspending the incredulity of them speaking in English about people you've never heard of, try to get some ideas about the speakers and the Zorks.

A. The Zorks are despicable – aggressive, brutal, and cruel!

B. Yeah!

You surely form an impression about the Zorks as aggressive brutes, but not only that, you would also get a picture of the conversants – probably neither A nor B thinks much of the Zorks.

Now, imagine another conversation.

C. The Zorks shot arrows, slashed with swords, and killed the Awds!

D. Yeah!

What do you think of the Zorks this time? You'd imagine that the Zorks were battling against the Awds. And what do you think C and D think of the Zorks? Perhaps C and D don't seem to think of the Zorks as negatively as A and B. By analyzing the choice of words and how things are described, you can begin to guess social relationships surrounding the speakers and the social world that they are talking about. A picture may be worth a thousand words, but words can speak a lot more than what they say.

Maass and her colleagues (e.g., Maass, 1999; Maass, Salvi, Arcuri, & Semin, 1989) have shown linguistic intergroup biases. People have biases in their language use to describe their outgroup's negative deeds with more abstract words (e.g., adjectives that describe dispositional charac-

teristics such as "aggressive" and "brutal"), but their ingroup's negative actions with more concrete words (e.g., verbs that describe actions such as "shot arrows" and "slashed with swords"). In contrast, positive deeds are more likely used for ingroups than for outgroups. So, the chances are that the Zorks are A and B's outgroup, but maybe C and D's ingroup – at least their friends if not their ingroup. The conversants engage in different linguistic practices (Holtgraves & Kashima, 2008) to share their views about their social reality – the social reality of their intergroup relations in particular – and reaffirming their shared reality (Echterhoff & Higgins, 2017; Hardin & Higgins, 1996) in the process.

More generally, communicators' choice of words can reveal their intersubjective social reality, that is to say, how the communicators understand and represent their social reality between them as they communicate with each other. In so doing, they maintain their intersubjective social reality by sharing information about the world in which it makes sense. As Berger and Luckmann (1966) noted more than half a century ago, "[m]ost conversation does not in so many words define the nature of the world. Rather, it takes place against the background of a world that is silently taken for granted (p. 172)." This chapter sketches out how everyday interpersonal communication may participate in the intersubjective construction and reconstruction of social reality as it drives the transmission of cultural information through social networks. In what follows, a model of cultural transmission is outlined, linked to information diffusion in social networks, and implications of this approach are explored.

Cultural Transmission

Conversations like the ones shown above are instances of cultural transmission. As people communicate with each other, they transmit cultural information, i.e., information that can be transmitted socially from one person to another and potentially influence a person's psychological processes, including cognition, affect, and behaviour. So, cultural information is distinguished from genetic information, which is also transmitted from one person to another, but only genetically and not socially (Kashima, Bain, & Perfors, 2019). The argument here is that cultural transmission, i.e., social transmission of cultural information, is critically involved in the construction of intergroup social reality (Kashima, 2014).

Kashima et al. (2019) proposed that cultural transmission involves at least four critical subprocesses: production, grounding, interpretation,

and memory. The first is production. As communicators communicate with their audience, they transform their cognitive representations stored in their memory into other forms that can be observed and interpreted by the audience. Whatever forms they may take – audible sounds, visible verbal or nonverbal signs, or other traces of behaviors such as pictures and memes – they can represent some cultural information. As we noted above, cultural information can pertain to the communicators' intergroup social reality, i.e., how communicators relate to the group of people about whom they are communicating. For example, research has examined how the communicators would tailor their messages for the audiences to whom the communications are directed. Higgins and Rholes (1978) showed that American university students tended to adjust their description of a person in a more positive (negative) direction if they thought their audience liked (disliked) the person, and they tended to remember what they described (for a review, see Echterhoff & Higgins, 2017).

Some of the outcomes of these communicative behaviors can become artefacts like leaflets, pamphlets and internet memes that circulate among people. However, the communications typically need to be grounded. Grounding means that the communicators establish their mutual, and intersubjective, understanding of the transmitted information (H. H. Clark, 1996; Kashima, Klein, & Clark, 2007). When someone presents a statement like "The Zorks are despicable" and someone else accepts it by saying, "Yeah!", this pair of statements can be seen as an indication that the information represented in the statement has been mutually understood and therefore grounded between them. The mutually understood information is then *interpreted* and stored in *memory* of those who are involved in the communication. This way, the communication process among people typically involves the externalization (production) and internalization (interpretation) of cultural information and the movement back and forth between the subjective and the intersubjective representations (Figure 1).
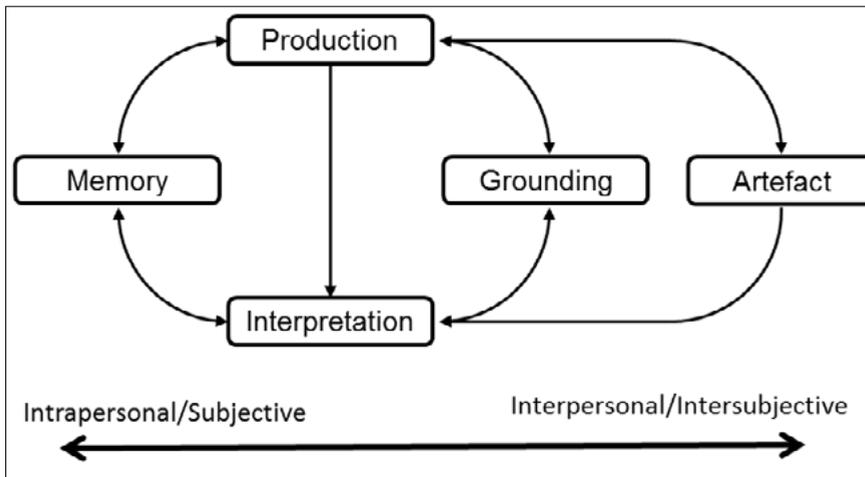
Figure 1: Subprocesses of cultural transmission (Based on Kashima et al., 2019).

Grounding is integral to our everyday communication, in which two or more individuals with similar levels of status, knowledge, and expertise are interacting with each other. Here communicators try to establish and build on their common ground – the set of information that the communicators actually share and believe that they share at the time (H. H. Clark, 1996; Kashima et al., 2007). Grounding is the process of adding the newly shared information to the communicators' common ground and further developing their shared understanding about the topic of their conversation. Furthermore, grounding is likely to sharpen the shared understanding about that which is discussed. On the one hand, grounded information is highlighted, rehearsed, and more likely deeply encoded than information that is not grounded. This would strengthen the memory traces of the grounded information. On the other hand, information that is not grounded is more likely forgotten.

This conjecture is based on Coman, HIrst, and their colleagues' work on retrieval induced forgetting (Coman & Hirst, 2012; Coman, Manier, & Hirst, 2009; Coman, Momennejad, Drach, & Geana, 2016; Cuc, Ozuru, Manier, & Hirst, 2006). For example, Coman et al. (2009) examined New Yorkers' memories about the September 11 attacks. They classified the participants' memories into categories such as TIME (e.g., I heard about them at 9am, I woke up at 8am that day), and recorded the conversation of randomly paired two participants about their experiences of the at-

tacks. Those who heard their conversation partner mention an event in one category (TIME – woke up at 7:30am) tended to retrieve their related, but unmentioned memory of that type (TIME – found out about the attacks at 9am) more slowly than unrelated and unmentioned memory (e.g., heard about them at home). Metaphorically speaking, grounding information is like putting a spotlight on the information. It highlights the information, but darkens its surrounding background, sharpening the contour of the memory of the grounded information.

Diffusion of Information on Social Networks

Cultural transmission occurs not just between two people or within a small group of interacting individuals, but in the broader context of social networks – patterns of social interaction that connect individuals and other actors in a society (e.g., Kashima, Bratanova, & Peters, 2017; Robins, 2015). Figure 2 depicts a social network that surrounds two individuals, A and B. If cultural transmission occurs from A to B, grounded cultural information may spread through the social networks from B to C to D and beyond into the community like a communication chain, or because A acts like a hub and repeatedly communicates the information to the acquaintances like E, F, and G.
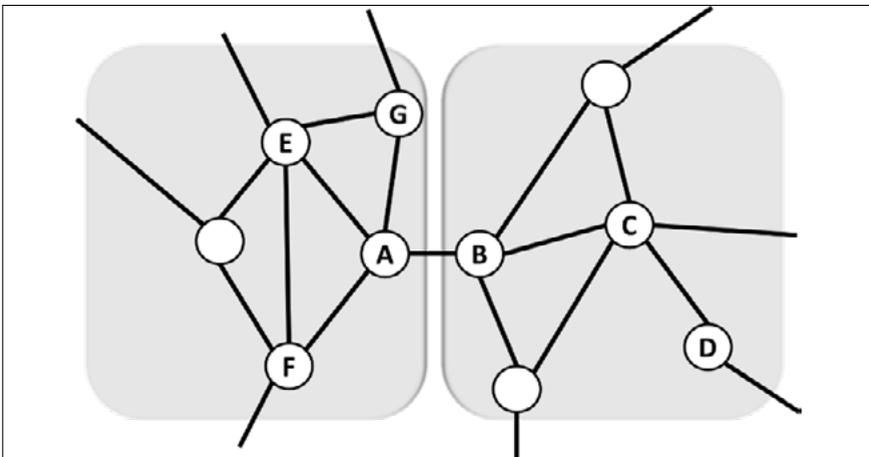
Figure 2. Cultural Transmission in Social Networks (Based on Kashima, Bratanova, & Peters, 2017)
Note. Circles represent nodes (actors) and lines, network ties indicating potentials for social interaction and information diffusion between the nodes. A-G are different actors in the network, and gray areas represent two different communities.

Cultural transmission on social networks can be simulated in laboratory experiments. A chain (e.g., A→B→C→D) can be simulated by the method of serial reproduction (Bartlett, 1932), in which information is given to the first person (A), who in turn communicates it from memory to the second person (B), who in turn communicates it for the third person (C), and so on. To simulate a "hub" (highly central node), the method of repeated reproduction (Bartlett, 1932) can be used, in which a participant repeatedly communicates cultural information to different audiences (e.g., A→B,E,F,G). Research using these methods has provided experimental evidence that sheer cultural transmission can contribute to the formation and maintenance of intersubjective social reality. A somewhat different, but related experimental paradigm was used by Jacobs and Campbell (1961) in their examination of transmission of cultural norms about autokinetic movements (also see Zucker, 1977).

Cultural transmission can form an intersubjective social reality in social networks

Imagine Adam and Ben (A and B in Figure 2) have a dispute. It starts off with an innocuous accident, Adam bumping into Ben in a narrow corridor, for example, and then Ben tripping over and spilling a coffee on Adam, etc. These unfortunate accidental innuendos get bigger and bigger as they begin to retaliate, escalating the conflict. The story can begin to circulate among Adam's and Ben's friends. What would happen as the story gets told and retold through a chain of communication?

Lee, Gelfand, and Kashima (2014) explored this question. Students were told about conflicts between two groups of university students, in which both groups performed equally blameworthy actions to the other side. In one condition, they took the perspective of a partisan (i.e., friends of one group or the other) and transmitted the story in serial reproduction chains; however, in the other condition, participants took the perspective of neutral observers in transmitting the story. Although the story remained evenly balanced with the neutral observers, and both groups were seen to be equally blameworthy, the partisan story became increasingly ingroup favoring and outgroup derogatory. Thus, sheer communication in an ingroup can turn an initially equally blameworthy intergroup conflict into an outgroup derogating prejudice, amplifying intergroup differentiation and forming an intersubjective social reality of intergroup conflict. Largely consistent with this, Thompson, Judd, and

Park (2000) also showed that stereotypes about a group acquired from multiple sources in social networks are likely to be more extreme and less likely to include information about the variability within the group (on communication and stereotypes, also see Brauer, Judd, & Jacquelin, 2001).

Sheer communication across multiple individuals can transform initially unstructured information into stereotype-like representations. Martin et al. (2014) made up "aliens" (e.g., blue, square, and bouncy) and attributed a random set of personality traits (e.g., cheerful, arrogant), thus creating alien-trait associations. They then had the participant to learn a subset of these alien-trait associations. The participant then reproduced those associations from memory by identifying which trait was associated with which alien. Those reproduced alien-trait associations were then used as stimuli for the second participant, who learned and reproduced them just like the first person. This procedure was repeated seven times. Martin et al.'s findings suggest that later reproductions were no longer random alien-trait associations, but reflected something more meaningful for the participants and easier to remember. For example, the participants in later parts of the serial reproduction chain could reproduce their preceding participants' reproductions more accurately. This is because alien-trait associations became more systematic and stereotype-like – the aliens with the same properties (e.g., color, shape, etc.) were attributed the same personality traits (e.g., blue squares are reliable and arrogant). In the end, stereotypes of alien categories were formed.

As information about social groups is shared among people through grounding in social networks, it can become their communal common ground – information collectively taken for granted within the group. This is akin to descriptive norms (e.g., Cialdini, Reno, & Kallgren, 1990), but it is something a little more than that. People not only share the information, but also perceive that they share it. It is the mutual understanding that multiple individuals share their understanding that matters (Shteynberg, Hirsh, Bentley, & Garthoff, 2020). Kashima, Wilson, Lusher, Pearson, and Pearson (2013) suggested that people can learn descriptive norms by observing what others do in their social networks, namely, through exemplar-based social category learning (Kashima, Woolcock, & Kashima, 2000). Perceptions of descriptive norms can facilitate a further spread of descriptions about the intersubjective social reality. Using the method of repeated reproduction, Bratanova and Kashima (2014) showed that a central node in a social network (i.e., a hub) can spread potentially biased information (e.g., prejudiced stereotypes) by repeating a biased

communication to multiple audiences when the speaker thought that it was consistent with the group's descriptive norm.

Cultural transmission can maintain intersubjective social reality in social networks

Intersubjective social reality tends to perpetuate itself. Put more theoretically, once certain ideas and practices about intergroup relations are in a group's collective common ground, cultural transmission tends to maintain the prevalent representations of the intergroup relations. An early demonstration came from Allport and Postman (1947). In a classical study on rumor transmission, they found that a story about a White man with a razor threatening a Black gentleman sometimes became a more stereotypical story of a Black man threatening a White man. Kashima (2000) more recently showed that stereotype consistent (SC) information is more likely to be transmitted than stereotype inconsistent (SI) information in serial reproduction chains. Here, the original narrative was about a man and a woman with gender SC and SI characteristics interacting with each other, performing gender SC and SI behaviours. Serial reproduction chains initially reproduced SI information more than SC in some cases, but eventually reproduced more SC than SI information in the end. This SC bias in serial reproduction is robust not only in written (A. E. Clark & Kashima, 2007; Hunzaker, 2014, 2016; Kashima, 2000; Lyons & Kashima, 2003), but also in oral communications (Kashima, Lyons, & Clark, 2013).

A. E. Clark and Kashima (2007; also Kashima et al., 2007) explained this in terms of the informativeness-connectivity dilemma. First of all, SI information is likely informative because it challenges the shared beliefs about the world and therefore has a chance of informing its recipient of the new social reality. However, it is also likely to be more difficult to communicate precisely because it challenges the commonsense (on a similar view of stereotypes, see Sherman, Lee, Bessenoff, & Frost, 1998). The audience may not only find it puzzling and hard to understand, but may ask for an explanation or even challenge the veracity of the information. This way, SI information may be disruptive to the communicators' interpersonal relationship. In contrast, SC information is likely socially connective because SC information "makes sense" against the taken-for-granted background that Berger and Luckmann (1966) talked about and in so doing affirms the collective common ground and the shared collective identity (Kashima et al., 2007). Thus, if the context of

communication demands informativeness, SI information may be communicated (e.g., Goodman, Webb, & Stewart, 2009; Lyons & Kashima, 2003). However, in default social contexts in which informativeness is secondary to social connectivity, communicators are likely to prefer to transmit socially connective information, namely, SC information.

Consistent with this, A. E. Clark & Kashima (2007) found that narrators of a story tended to regard SI information in the story as more informative, but less socially connective than SC information. They also saw SC information as more communicable – easy to communicate – than SI information. Consequently, the narrators were more likely to mention SC than SI in serial reproduction chains. It is important to note that SC information is seen to be more socially connective than SI information because the stereotype is seen to be shared and endorsed with others. This means that if the stereotype is not upheld by others, the foundational basis of the SC bias is undermined. Indeed, when participants were led to believe that the stereotype was not endorsed by others in the group, an SC bias disappeared. Kurz and Lyons's (2009) findings corroborate this line of reasoning. They found that an SC bias was most pronounced when people were communicating about an outgroup member to their ingroup members.

This is not to say that people knowingly transmit erroneous stereotypical information. People tend to transmit what they believe is valid and informative (Lyons & Kashima, 2003), especially when there is information asymmetry, that is, when the sender does, but receivers do not, have given information, and that they both know that there is this information asymmetry. Lyons and Kashima (2003) showed that people tried to be informative and communicated SI information when they knew with certainty that their audience had known the stereotype but did not have the SI information. Intriguingly, when people knew with certainty that their audience did not know the stereotype (i.e., ignorant audience), they communicated SC information more as if they were trying to "teach" the stereotype to the ignorant audience. This is analogous to the situation in which adults are "teaching" their ingroup children what their outgroup is like. Additional analysis suggested that this was driven by the narrators' motivation to convey truthful information. Thus, cultural information that reflects intersubjective social reality may be transmitted across generations into the future.

Summary

Information about intergroup relations can be communicated in the form of stereotypes about a group. As it spreads on social networks, it can become shared and established as part of the network's communal common ground. Through the process of grounding, the shared information about the intergroup relations can become their intersubjective social reality. Once established as such, intersubjective social reality can perpetuate itself as people engage in their everyday conversations. As Maass (1999) noted more than two decades ago, "biased language use tends to feed into a cycle in which initial beliefs such as stereotypes are protected against disconfirmation, which in turn contributes to the stability of intrapersonally as well as interpersonally shared beliefs systems (p. 105)." Her insights have survived the test of time, and will remain true as a guiding light that shines into the future.

Concluding Comments

The Twin Earthians' conversation at the start of the chapter is just one of many that can take place every minute in face to face social interactions. However, technological advances can further enhance the capacities of intelligent beings to communicate in a virtual reality. Looking at the real Earth, globalizing humanity is communicating at a much greater scale and a faster pace than before. George Floyd, Greta Thunberg, and Mahsa Amini are all household names as I write this chapter in 2022. Their deeds and plights are shown repeatedly on traditional mass media as well as social media sites, with an increasing emphasis on the latter. Comments are left and messages are sent. Tweets and conversations can happen on the internet in real time between people who are physically located half a planet away. In some cases, those contributing to the discourse do not have to be people in flesh, but artificial intelligence in silico.

What may be the implication of the cultural transmission processes in this hyperconnected world? On the one hand, enhanced capabilities for social interaction can potentially increase the chance of people having productive conversations. People can not only socialize, but also engage in conversations, discussions, and even civilized debates about matters that are important to many. Indeed, some cautious optimism was expressed by some that the internet can facilitate deliberative democracy – reasoned discourse among people to voice their opinions and discuss

the matters of public interest (e.g., Blumler & Gurevitch, 2001; Gimmler, 2001; Papacharissi, 2002). On the other hand, the same capabilities can increase the chance of like-minded individuals finding each other and sorting themselves into fragmented and closed clusters of individuals (e.g., Dahlberg, 2007; Dahlgren, 2005; Sunstein, 2018; Wilhelm, 2002), further driving each other's opinions in the more extreme direction (Isenberg, 1986; Myers & Lamm, 1976), and forming opinion-based groups (McGarty, Bliuc, Thomas, & Bongiorno, 2009).

Put simply, hyperconnected interactions can integrate or fragment society. Without a doubt, broad cultural, institutional, and technological contexts can play a significant role in determining which way our society might go. However, the current perspective suggests that our everyday conversations online and offline can be critical. Depending on how people characterize others by choice of words and phrases, they can discursively construct and reconstruct intersubjective social reality of societal integration or fragmentation. Perhaps by characterizing those who disagree with us with abstract words, we may be conveying the picture of them as an objectified social entity that fragments society; by agreeing to such characterizations, we may be affirming and helping construct the intersubjective social reality of fragmentation. People may then develop the perception of structural differentiation within the society – the view that the society is fragmented into subgroups that hold irreconcilable worldviews (Koudenburg & Kashima, 2022). Koudenburg and Kashima's work suggests that opinion differentiation – that people have diverse opinions, pro and con, on an issue – in and of itself does not discourage, and in fact even encourage, discussion. Rather, it is the perception of structural differentiation that stifles potentially constructive discussions about controversial issues. This can further exacerbate societal fragmentation.

As the 21st century deepens, humanity is confronted by many challenges. Not only climate change linked extreme weather events can disrupt the lives and livelihoods of hundreds of thousands of people, but geopolitical tensions, conflicts, and warfare can threaten the integrity of our society. Faced with these challenges, it is not too surprising that we are engaging in massive global conversations about our collective futures. Whether the hyperconnectivity afforded by the technological advances can help globalizing humanity to meet the challenges, or hinder the collective effort, may in part depend on how we talk to each other. If people can develop linguistic practices that retain opinion diversity without structural differentiation, our everyday conversations may be able to

constructively engage with a variety of societal issues that confront us in the 21st century.

References

Allport, G. W., & Postman, L. (1947). *The psychology of rumor*. Oxford, England: Henry Holt.

Bartlett, F. c. (1932). *Remembering: A study in experimental and social psychology*. Cambridge, UK: Cambridge University Press.

Berger, P., & Luckmann, T. (1966). *The social construction of reality*. London, UK: Penguin Books.

Blumler, J. G., & Gurevitch, M. (2001). The new media and our political communication dicontents: Democratizing cyberspace. *Information, Communication & Society, 4*(1), 1-13. doi:10.1080/713768514

Bratanova, B., & Kashima, Y. (2014). The "saying is repeating" effect: dyadic communication can generate cultural stereotypes. *J Soc Psychol, 154*(2), 155-174. doi:10.1080/00224545.2013.874326

Brauer, M., Judd, C. M., & Jacquelin, V. (2001). The communication of social stereotypes: The effects of group discussion and information distribution on stereotypic appraisals. *Journal of Personality and Social Psychology, 81*(463-475).

Cialdini, R. B., Reno, R. R., & Kallgren, C. A. (1990). A focus theory of normative conduct: Rrecycling the concept of norms to reduce littering in public places. . *Journal of Personality and Social Psychology, 58*(6), 1015-1026.

Clark, A. E., & Kashima, Y. (2007). Stereotypes help people connect with others in the community: A situated functional analysis of the stereotype consistency bias in communication. *Journal of Personality and Social Psychology, 93*(6), 1028-1039. doi:10.1037/0022-3514.93.6.1028

Clark, H. H. (1996). *Using language*. New York, NY: Cambridge University Press.

Coman, A., & Hirst, W. (2012). Cognition through a social network: The propagation of induced forgetting and practice effects. *Journal of Experimental Psychology: General, 141*(2), 321-336. doi:10.1037/a0025247

Coman, A., Manier, D., & Hirst, W. (2009). Forgetting the unforgettable through conversation: Socially shared retrieval-induced forgetting of September 11 memories. *Psychological Science, 20*(5), 627-633.

Coman, A., Momennejad, I., Drach, R. D., & Geana, A. (2016). Mnemonic convergence in social networks: The emergent properties of cognition at a collective level. *Proceedings of the National Academy of Sciences,*

201525569.

Cuc, A., Ozuru, Y., Manier, D., & Hirst, W. (2006). On the formation of collective memories: The role of a dominant narrator. *Mem Cognit, 34*(4), 752-762. doi:10.3758/bf03193423

Dahlberg, L. (2007). Rethinking the fragmentation of the cyberpublic: From consensus to contestation. *New Media & Society, 9*(5), 827-847. doi:10.1177/1461444807081228

Dahlgren, P. (2005). The Internet, public spheres, and political communication: Dispersion and deliberation. *Political Communication, 22*(2), 147-162.

Echterhoff, G., & Higgins, E. T. (2017). Creating shared reality in interpersonal and intergroup communication: the role of epistemic processes and their interplay. *European Review of Social Psychology, 28*(1), 175-226. doi:10.1080/10463283.2017.1333315

Gimmler, A. (2001). Deliberative democracy, the public sphere and the internet. *Philosophy & Social Criticism, 27*(4), 21-39.

Goodman, R. L., Webb, T. L., & Stewart, A. J. (2009). Communicating stereotype-relevant information: is factual information subject to the same communication biases as fictional information? *Pers Soc Psychol Bull, 35*(7), 836-852. doi:10.1177/0146167209334780

Hardin, C. D., & Higgins, E. T. (1996). Shared reality: How social verification makes the subjective objective. In R. M. Sorrentino & E. T. Higgins (Eds.), *Handbook of motivation and cognition: The interpersonal context* (Vol. 3, pp. 28-84). New York, NY: Guilford Press.

Higgins, E. T., & Rholes, W. S. (1978). "Saying is believing": Effects of message modification on memory and liking for the person described. *Journal of Experimental Social Psychology, 14*(4), 363-378. doi:https://doi.org/10.1016/0022-1031(78)90032-X

Holtgraves, T. M., & Kashima, Y. (2008). Language, meaning, and social cognition. *Pers Soc Psychol Rev, 12*(1), 73-94. doi:10.1177/1088868307309605

Hunzaker, M. B. F. (2014). Making sense of misfortune: Cultural schemas, victim redefinition, and the perpetuation of stereotypes. *Social Psychology Quarterly, 77*(2), 166-184. doi:10.1177/0190272514521219

Hunzaker, M. B. F. (2016). Cultural Sentiments and Schema-Consistency Bias in Information Transmission. *American Sociological Review, 81*(6), 1223-1250. doi:10.1177/0003122416671742

Isenberg, D. J. (1986). Group polarization: A critical review and meta-analysis. *Journal of Personality and Social Psychology, 50*(6), 1141.

Jacobs, R. C., & Campbell, D. T. (1961). The perpetuation of an arbitrary

tradition through several generations of a laboratory microculture. *The Journal of Abnormal and Social Psychology, 62*(3), 649-658. doi:10.1037/h0044182

Kashima, Y. (2000). Maintaining cultural stereotypes in the serial reproduction of narratives. *Personality and Social Psychology Bulletin, 26*(5), 594-604. doi:doi.org/10.1177/0146167200267007

Kashima, Y. (2014). Meaning, grounding, and the construction of social reality. *Asian Journal of Social Psychology, 17*(2), 81-95. doi:10.1111/ajsp.12051

Kashima, Y., Bain, P., & Perfors, A. (2019). The psychology of cultural dynamics: What is it, what is known, and what is yet to be known? . *Annual Review of Psychology, 70*, 499-529. doi:https://doi.org/10.1146/annurev-psych-010418-103112

Kashima, Y., Bratanova, B., & Peters, K. (2017). Social Transmission and Shared Reality in Cultural Dynamics. *Current Opinion in Psychology*.

Kashima, Y., Klein, O., & Clark, A. E. (2007). Grounding: Sharing information in social interaction. In K. Fiedler (Ed.), *Social communication* (pp. 27-77). New York, NY: Psychology Press.

Kashima, Y., Lyons, A., & Clark, A. (2013). The maintenance of cultural stereotypes in the conversational retelling of narratives. *Asian Journal of Social Psychology, 16*(1), 60-70. doi:10.1111/ajsp.12004

Kashima, Y., Wilson, S., Lusher, D., Pearson, L. J., & Pearson, C. (2013). The acquisition of perceived descriptive norms as social category learning in social networks. *Social Networks, 35*, 711-719. doi:10.1016/j.socnet.2013.06.002

Kashima, Y., Woolcock, J., & Kashima, E. S. (2000). Group impressions as dynamic configurations: The tensor product model of group impression formation and change. *Psychological Review, 107*(4), 914-942.

Koudenburg, N., & Kashima, Y. (2022). A polarized discourse: Effects of opinion differentiation and structural differentiation on communication. *Personality and Social Psychology Bulletin, 48*(7), 1068-1086. doi:https://doi.org/10.1177/01461672211030816

Kurz, T., & Lyons, A. (2009). Integroup influences on the stereotype consistency bias in communication: Does it matter who we are communication about and to whom we are communicating? *Social Cognition, 27*, 893-904.

Lee, T. L., Gelfand, M. J., & Kashima, Y. (2014). The serial reproduction of conflict: Third parties escalate conflict through communication biases. *Journal of Experimental Social Psychology, 54*(1), 68-72. doi:10.1016/j.jesp.2014.04.006

Lyons, A., & Kashima, Y. (2003). How are stereotypes maintained through communication? The influence of stereotype sharedness. *Journal of Personality and Social Psychology, 85*(6), 989-1005. doi:10.1037/0022-3514.85.6.989

Maass, A. (1999). Linguistic intergroup bias: Stereotype perpetuation through language. In *Advances in Experimental Social Psychology* (Vol. 31, pp. 79-121): Elsevier.

Maass, A., Salvi, D., Arcuri, L., & Semin, G. R. (1989). Language use in intergroup contexts: The linguistic intergroup bias. *Journal of Personality and Social Psychology, 57*(6), 981-993.

Martin, D., Hutchison, J., Slessor, G., Urquhart, J., Cunningham, S. J., & Smith, K. (2014). The spontaneous formation of stereotypes via cumulative cultural evolution. *Psychological Science, 25*(9), 1777-1786. doi:10.1177/0956797614541129

McGarty, C., Bliuc, A. M., Thomas, E. F., & Bongiorno, R. (2009). Collective action as the material expression of opinion-based group membership. *Journal of Social Issues, 65*(4), 839-857. doi:doi.org/10.1111/j.1540-4560.2009.01627.x

Myers, D. G., & Lamm, H. (1976). The group polarization phenomenon. *Psychological Bulletin, 83*(4), 602.

Papacharissi, Z. (2002). The virtual sphere:The internet as a public sphere. *New Media & Society, 4*(1), 9-27. doi:10.1177/14614440222226244

Robins, G. (2015). *Doing Social Network Research: Network-based Research Design for Social Scientists*. London: SAGE.

Sherman, J. W., Lee, A. Y., Bessenoff, G. R., & Frost, L. A. (1998). Stereotype efficiency reconsidered: Encoding flexibility under cognitive load. *Journal of Personality and Social Psychology, 75*(3), 589-606.

Shteynberg, G., Hirsh, J. B., Bentley, R. A., & Garthoff, J. (2020). Shared worlds and shared minds: A theory of collective learning and a psychology of common knowledge. *Psychological Review, 127*(5), 918-931. doi:10.1037/rev0000200

Sunstein, C. R. (2018). *#Republic*: Princeton University Press.

Thompson, M. S., Judd, C. M., & Park, B. (2000). The Consequences of Communicating Social Stereotypes. *Journal of Experimental Social Psychology, 36*(6), 567-599. doi:10.1006/jesp.1999.1419

Wilhelm, A. G. (2002). *Democracy in the digital age: Challenges to political life in cyberspace*: Routledge.

Zucker, L. G. (1977). The role of institutionalization in cultural persistence. *American Journal of Sociology, 42*(5), 726-743. doi:10.2307/2094862

2. Fiction and Social Cognition

Emanuele Castano
University of Florence, Italy

In the Beginning Was the Stories

The development of language in homo sapiens has allowed for an exponential growth in our ability to systematize knowledge, accumulate it and transmit it from one generation to the next, thus providing great evolutionary advantages to our species. Language is however not to be regarded as merely a code that we use for such purposes. It is a constitutive and integral part of who we are. Researchers have suggested that social communication, which for homo sapiens is tantamount to language, and particularly complex language, may have developed as a response to an evolutionary pressure for the development of more sophisticated social cognition to sustain sociability and social functioning. On the other hand, it also acknowledged that once in place, language may have allowed for the development of, and further scaffolded, increasingly sophisticated social cognition (Oesch & Dunbar, 2017).

Anthropological and cognitive science theories suggest, moreover, that language, and especially stories, have made a fundamental contribution to the development of our social organization, helping us to create cohesion in the small groups in which our ancestors lived as early as a million years ago. This had important repercussions on our capacity to cooperate, probably contributing significantly to our hypersociality (Tomasello, 2014).

One of the key factors in the development of such hyper-sociality is the discovery and especially the consequences of our ancestors acquiring

control of fire, the latter dating approximatively 300,000 years ago. Being able to light a fire led to great developments. For example, it allowed us to cook the meat of captured animals, with important effects on our morphology, particularly that of our brains, due to improvements in the efficiency in metabolizing (Wrangham, 2009). For our present concerns, however, an important development is the fact that after sunset, gathered around the hearth on a dark planet (and perhaps inspired by a starry sky), our ancestors began to tell stories to each other. These are the reflections made of anthropologist Polly Wiessner, based on her observations of the life of the African Ju'hoan (ung) tribe: "Nighttime activities reduce the tensions of the day through singing, dancing, religious ceremonies, and compelling stories, often about familiar people. Such stories describe the functioning of entire institutions in a small-scale society with little formal teaching. Nightly discourses play an important role in evoking higher orders of theory of mind through imagination, conveying the attributes of people within broad networks (virtual communities) as well as the "big picture" of cultural institutions that generate regularities of behavior, cooperation, and trust at the regional level." (Wiessner, 2014)

Stories and Social Cognition

The concept of Theory of Mind, mentioned by Wiessner, is commonly defined in the psychological sciences as the ability to infer others' mental processes, their emotions, beliefs and aspirations, their thoughts, doubts, desires. In other words, to understand the mind of others. This ability, not entirely absent in other primates and perhaps other mammals, certainly finds its most advanced form in homo sapiens. In fact, the work I briefly referred to above concerning the relationship between language and social cognition (Oesch & Dunbar, 2017), has focused primarily on Theory of Mind as an example of social cognition. Such a capacity is key in making our hypersociality possible, and, in turn, the pressure to maintain and develop such hypersociality pushes us to refine our Theory of Mind. Wiessner, however, advances a more precise hypothesis about the role of narrative, of stories. Is she right? Does reading stories teach us to read the minds of others? It would seem so.

A couple of decades ago, Raymond Mar, Keith Oatley and their colleagues investigated the relationship between engagement with stories, specifically, fiction, and Theory of Mind (Mar et al., 2006). Exposure to fiction was measured using an adapted version of the Author Recognition

Test, which consists of a list of names of fiction and nonfiction authors. Participants were asked to select the names of authors they knew, then allowing the researchers to calculate, for each participant, an exposure score to fiction and nonfiction. Next, participants performed a series of tests, including the Reading the Mind in the Eyes (RMET; Baron-Cohen et al., 2001). The RMET, among the most widely used Theory of Mind measures, allows to assess accuracy in inferring others' mental states by observing only the eye region. Results showed that while exposure to non-fiction did not correlate with performance on the RMET, exposure to fiction did: the more exposure, the greater the performance on the RMET. These results can be interpreted as evidence that fiction is a simulation of social life: readers are transported into the imaginary world, identify with the characters, feel their emotions, and imagine their thoughts and desires. In this way, they hone their mind-reading skills (Oatley, 1999). Recent research has, however, qualified this conclusion and proposed a different account of the effect of fiction on mind reading.

Types of Fiction, Theory of Mind, Theory of Society

A decade ago, I conducted with my then-graduate student, David Kidd, a series of experiments aimed at more directly testing the potential causal effect of reading fiction on ToM (Kidd & Castano, 2013). Participants were asked to read one among many excerpts from novels or short stories previously classified as literary or popular fiction. Assignment to the specific reading was of course random. Next, the participants completed the RMET and other tasks meant to measure Theory of Mind. Results showed only those who had read literary fiction scored higher on ToM test; those who had read popular fiction scored similarly to participants who had read nothing or who had read nonfiction. This result has been replicated using either the exact same methodology (e.g., Kidd & Castano, 2019; Kidd et al., 2016; van Kuijk et al., 2018) or conceptually similar ones (Black & Barnes, 2015; Castano, 2021; Pino & Mazza, 2016; Schwerin & Lenhart, 2022). Some studies did not replicate this finding (e.g., Panero et al., 2016), but their methodology has been questioned (Kidd & Castano, 2017a). Reviews of the experimental work on the effects of brief exposure to fiction on ToM have concluded that the effect of reading fiction on ToM is reliable, if small, but they have not focused on the specific distinction between literary and popular fiction (e.g., Dodell-Feder & Tamir, 2018; Quinlan et al., 2022).

The differential impact of exposure to literary versus popular fiction is also supported by correlational research. As mentioned, the relation between exposure to fiction and ToM was first investigated using the Author Recognition Test. In studies such as the pioneer one by Mar et al. (2006), a score reflecting the degree of exposure to fiction is given to each participant, based on his or her responses to the Author Recognition Test. In more recent studies, however, instead of a single score, two scores are calculated: one for exposure to popular fiction and one for exposure to literary fiction (Kidd & Castano, 2017b). Confirming the results obtained with the experimental methodology, these studies revealed that while exposure to popular fiction is not associated with performance on Theory of Mind tasks, exposure to literary fiction is: the greater the exposure to literary fiction, the better Theory of Mind (see also Castano et al., 2020; 2021; Das & Vasudevan, 2022).

With the important caveat of what type of fiction is under consideration, research findings overall support the conjecture by Wiessner (2014). Theory of Mind, however, is only one of the aspects mentioned by Wiessner in her analysis of the impact of storytelling on human societies. Do stories do other things as well? Do they also serve to "convey the attributes of people within broad networks (virtual communities) as well as the 'general framework' of cultural institutions…?" A tentative answer to this question is, yes: In addition to Theory of Mind, reading narrative fiction may also impact on our Theory of Society. Theory of Society (ToS) can be defined as a special purpose modular capacity to understand others in terms of culturally transmitted information about group membership, for example, which social groups exist in one's culture and which stereotypes adhere to these groups (Hirschfeld et al., 2007; p. 451; Hirschfeld, 2006). Contrary to ToM, the construct of ToS has elicited, strictly speaking, virtually no empirical research. Yet, I would argue that this broad construct covers a host of social psychological constructs, among which are psychological essentialism, the belief that some categories possess an underlying essence common to all members and causally responsible for their typical attributes and behaviors (see Medin & Atran, 2004), and attributional complexity, the tendency to see others' behavior as resulting from complex interactions of internal and external causes, attempt to integrate seemingly inconsistent information about others, and reflect on their own insight into others' behavior (Fletcher et al., 1986). Both of these constructs have been studied in relation to fiction exposure, with findings showing that exposure to literary fiction is associated with

lower psychological essentialism and higher attributional complexity. Exposure to popular fiction, on the other hand, is associated with lower attributional complexity and, to a lesser extent, greater psychological essentialism (Castano et al., 2020; 2021; see also Buttrick et al., 2022).

Against a hierarchy of fiction

If we look at the two sets of findings, which I have organized, primarily for aesthetics purposes, into Theory of Mind vs. Theory of Society, a parallel emerge with a distinction that is well-known to social psychologists, namely between individuated- vs. category-based perception (e.g., Brewer, 1988; Fiske & Neuberg, 1990), as well as the most recently proposed distinction in the literature on morality, between individuating vs. binding moral foundations (Haidt, 2002). Some of the socio-cognitive processes and cognitive styles discussed above, such as attributing an essence to social categories, facilitate the binding among humans that is functional to forming and maintaining social groups by allowing us to imagine (Anderson, 1983), constitute and maintain social groups, notably through the process of self and social categorization (Turner et al., 1987) and the development of social identities (Tajfel, 1981). Other socio cognitive processes and cognitive styles, such as the enhanced attention and accuracy in inferring people's thoughts and feelings, as well as interpreting the complexity of human affairs and rejecting essentialistic views of social categories, are individuating: They foster a view of the world in terms of unique individuals, as opposed to social groups. The research reviewed here seems thus to indicate that literary and popular fiction foster, primarily, individuating versus binding processes, respectively. Social psychology has traditionally viewed the former processes as providing the socio-cognitive substrate for the enactment of ethical, just behavior; and the latter, conceived either as cognitive shortcuts or as processes motivated by the need to maintain collective self-esteem and the social order, as underlying negative social behavior such as stereotyping and prejudice. If the former are fostered by literary, and the latter by popular, fiction, we might be tempted to conclude in favor of the hierarchy of fiction, which parallels the well-known distinction between highbrow and lowbrow fiction, respectively. This may be misguided. For a human society to have enough cohesion to function but also continue to evolve, the centripetal, aggregating action of popular fiction is just as important as the disruptive, centrifugal action of literary fiction. The need for an equi-

librium at the societal level finds also a parallel at the intra-individual level, with research showing the importance of a balance between a need for uniqueness and inclusion for psychological equanimity (Brewer, 1991).

In favor of a taxonomy of fiction

It is not uncommon in the social sciences, and even more so in society at large, to deny differences in the process of arguing for equality. Having just argued against a hierarchy of fiction, should I also deny that they differ? I will do the opposite.

The debate about the nature of a text, and thus also of the nature of fiction in particular, has a long history in disciplines such as literary studies, semiotics and linguistics. When I first started this line of inquiry, I was influenced primarily by Roland Barthes, whom on the one hand distinguished between writerly and readerly fiction (Kidd & Castano, 2013), and, on the other, argued for the need to shift the attention away from characteristics of the text, and redirect such attention onto the readers and their subjective experience. This view is known as the reader-response approach to literature; an approach which unifies scholars who stress the role of the reader in creating meaning through their reading of the text (e.g. Bleich, 1975, 1978; Holland, 1968, 1973, 1975; Iser, 1974, 1978). The reader-response approach developed as a criticism of earlier formalists theories in literary studies, which largely (and purposely) ignore factors such as cultural contexts, the intentions of the author of the text, and even the content of the text, and focus instead on the text itself (Richter, 2007; Tyson, 1999; Habib, 2011). Readers play an active role in constructing meaning (e.g. Bruner, 1987; Gerrig, 1993; Sharma, 2020; Miall & Kuiken, 1994), individual differences moderate readers' interest in and experience of the text (e.g. Rain & Mar, 2021), and the taxonomy of text, and particularly fiction, is influenced by historical, cultural and broad socio-political factors (McGregor, 1997; Gans, 1974; Ross, 1989; Walter, 1968). Yet, I would argue that if exposure to literary vs. popular fiction has differential effects on our social cognition, some difference must exist between the two texts; and since they are texts, the difference must reside in the language they use. A first support of this idea comes from a study comparing a corpus of English-language literary and popular fiction from the last two decade, in which we found that literary fiction has greater lexical and syntax complexity than popular fiction (Castano et al., 2024). This finding is consistent with the claim that literary fiction is

more complex than popular fiction (Kidd & Castano, 2013) and dovetails with evidence of a relationship between syntax complexity and Theory of Mind (Oesch & Dunbar, 2017).

A conclusive thought

In this brief chapter I have summarized ideas and research findings that my former students, colleagues and collaborators have developed and gathered over the last decade. In a funny turn of events, some 20 years after she had been my professor, and precisely 10 years ago, I had invited Anne Maass as a visiting professor in my department. At that time, many of the ideas I discussed here, were in an embryonic state. Anne being among the few social psychologists who had paid attention to language since the 1980s (Maass et al., 1989; 1996), we brainstormed about the language of popular and literary fiction, and she was instrumental in helping us formulating the hypothesis that one of the ways in which literary fiction may elicit the effects that we were beginning to observe, was through the use of more concrete language. It did take me 10 years to get around to test this hypothesis, but as I finish this chapter I will start to write an article in which data are presented supporting precisely such a conjecture.

References

Anderson, D. (1983). Imagined communities: Reflections on the origin and spread of nationalism. London: Verso.

Baron-Cohen, S., Wheelwright, S., Hill, J., Raste, Y., & Plumb, I. (2001). The "Reading the Mind in the Eyes" Test revised version: A study with normal adults, and adults with Asperger syndrome or high-functioning autism. *Journal of Child Psychology and Psychiatry*, 42, 241–251. doi:10.1111/1469-7610.00715.

Baron-Cohen, S., Wheelwright, S., Hill, J., Raste, Y., & Plumb, I. (2001). The "Reading the Mind in the Eyes" Test revised version: A study with normal adults, and adults with Asperger syndrome or high-functioning autism. *Journal of Child Psychology and Psychiatry*, 42, 241–251. doi:10.1111/1469-7610.00715.

Black, J., & Barnes, J. (2015). The effects of reading material on social and non-social cognition. Poetics, 52, 32–43. doi:10.1016/j.poetic.2015.07.001

Bleich, D. (1975). Readings and feelings: An introduction to subjective criticism. Urbana: National Council of Teachers of English.

Brewer, M. B. (1988). A dual process model of impression formation. In R. S. Wyer Jr., & T. K. Srull (Eds.), Advances in social cognition (Vol. 1, pp. 1–36). Hillsdale, NJ: Erlbaum.

Brewer, M. B. (1991). The social self: On being the same and different at the same time. *Personality and Social Psychology Bulletin*, 17(5), 475-482.

Bruner, J. (1987). Life as narrative. Social Research, 54(1), 11-32.

Buttrick, N., Westgate, E. C., & Oishi, S. (2022). Reading literary fiction is associated with a more complex worldview. *Personality and Social Psychology Bulletin*, 01461672221106059.

Castano, E. (2021). Art films foster theory of mind. Humanities and Social Sciences Communications, 8(1), 1-10.

Castano, E., Martingano, A. J., Perconti, P. (2020). The effect of exposure to fiction on attributional complexity, egocentric bias and accuracy in social perception. PLoS ONE, 15(5), e0233378.

Castano, E., Paladino, M. P., Cadwell, O. G., Cuccio, V., & Perconti, P. (2021). Exposure to literary fiction is associated with lower psychological essentialism. Frontiers in Psychology, 12, 662940. doi:10.3389/fpsyg.2021.662940

Castano, E., Zanella, J., Saedi, F., Zunshine, L., & Ducceschi, L. (2024). On the complexity of literary and popular fiction. Empirical Studies of the Arts, 42(1), 281-300.

Das, A., & Vasudevan, M. H. (2022). Reading patterns, engagement style and theory of mind. Psychological Studies, 1-9.

Dodell-Feder, D., & Tamir, D. I. (2018). Fiction reading has a small positive impact on social cognition: A meta-analysis. Journal of Experimental Psychology: General, 147(11), 1713–1727. https://doi.org/10.1037/xge0000395

Fiske, S. T., & Neuberg, S. L. (1990). A continuum of impression formation, from category-based to individuating processes: Influences of information and motivation on attention and interpretation. Advances in Experimental Social Psychology, 23, 1-74.

Gans, H. J. (1974). Popular culture and high culture: An analysis and evaluation of taste. Basic Books.

Gerrig, R. J. (1993). Experiencing narrative worlds: On the psychology activities of reading. Yale University Press.

Habib, M. A. R. (2011). Literary criticism from Plato to the present: An introduction. Oxford: Blackwell.

Hirschfeld, L. A. (2006). Who needs a theory of mind? In R. Viale, D. Andler, & L. Hirschfeld (Eds.), Biological and cultural bases of human inference (131-160). Mahwah, New Jersey: Lawrence Erlbaum.

Hirschfeld, L., Bartmess, E., White, S., & Frith, U. (2007). Can autistic children predict behavior by social stereotypes? Current Biology, 17(12), R451-R452.

Holland, N. (1968). The dynamics of literary response. New York: Oxford University Press.

Holland, N. (1973). Poems in persons. New York: Norton.

Holland, N. (1975). Readers reading. New Haven, CT: Yale University Press.

Iser, W. (1974). The implied reader: Patterns of communication in prose fiction from Bunyan to Beckett. Baltimore: Johns Hopkins University Press.

Iser, W. (1978). The act of reading: A theory of aesthetic response. Baltimore: Johns Hopkins University Press.

Kidd, D. C., & Castano, E. (2013). Reading literary fiction improves theory of mind. Science, 342, 377–380.

Kidd, D. C., & Castano, E. (2017a). Panero et al. (2016): Failure to replicate methods caused the failure to replicate results. Journal of Personality and Social Psychology, 112 (3), e1–e4

Kidd, D. C., & Castano, E. (2017b). Different stories: How levels of familiarity with literary and genre fiction relate to mentalizing. Psychology of Aesthetics, Creativity, and the Arts, 11(4), 474–486. https://doi.org/10.1037/aca0000069

Kidd, D. C., & Castano, E. (2019). Reading literary fiction and theory of mind: Three pre-registered replications and extensions of Kidd and Castano (2013). Social Psychological and Personality Science, 10, 522–531.

Kidd, D., Ongis, M., & Castano, E. (2016). On literary fiction and its effects on theory of mind. Scientific Study of Literature, 6, 42–58. doi:10.1075/ssol.6.1.04kid

Maass, A., Salvi, D., Arcuri, L., & Semin, G. R. (1989). Language use in intergroup contexts: The linguistic intergroup bias. Journal of personality and social psychology, 57(6), 981.

Mar, R. A., Oatley, K., Hirsh, J., dela Paz, J., & Peterson, J. B. (2006). Bookworms versus nerds: Exposure to fiction versus non-fiction, divergent associations with social ability, and the simulation of fictional social worlds. Journal of Research in Personality, 40, 694–712.

McGregor, C. (1997). Class in Australia (1 ed.). Penguin Books Australia

Ltd.

Medin, D. L., & Atran, S. (2004). The native mind: biological categorization and reasoning in development and across cultures. Psychological Review, 111(4), 960.

Miall, D. S., & Kuiken, D. (1994). Foregrounding, defamiliarization, and affect: Response to literary stories. Poetics, 22(5), 389-407.

Oatley, K. (1999). Why fiction may be twice as true as fact: Fiction as cognitive and emotional simulation. Review of General Psychology, 3(2), 101-117.

Oesch, N., & Dunbar, R. I. (2017). The emergence of recursion in human language: mentalizing predicts recursive syntax task performance. Journal of Neurolinguistics, 43, 95-106.

Panero, M. E., Weisberg, D. S., Black, J., Goldstein, T. R., Barnes, J. L., Brownell, H., & Winner, E. (2016). Does reading a single passage of literary fiction really improve theory of mind? An attempt at replication. Journal of Personality and Social Psychology, 111(5), e46.

Quinlan, J. A., Padgett, J. K., Khajehnassiri, A., & Mar, R. A. (2022). Does a brief exposure to literary fiction improve social ability? Assessing the evidential value of published studies with a p-curve. Journal of Experimental Psychology: General.

Rain, M., & Mar, R. A. (2021). Adult attachment and engagement with fictional characters. Journal of Social and Personal Relationships. https://doi.org/10.1177/02654075211018513

Richter, D. H. (2007). The critical tradition: Classic texts and contemporary trends. Boston: Bedford/St. Martin's.

Ross, A. (1989). No respect: Intellectuals & popular culture. Routledge.

Schwerin, J., & Lenhart, J. (2022). The effects of literariness on social-cognitive skills: Examining narrative engagement, transportation, and identification as moderators. Psychology of Aesthetics, Creativity, and the Arts. https://doi.org/10.1037/aca0000514

Sharma, N. (2020). Deriving meanings out of a fictional text: Analyzing readers' performance of a narrative in India by using a mental models approach. Participations: Journal of Audience and Reception Studies, 17(1), 67–91.

Tajfel, H. (1981). Human groups and social categories: Studies in social psychology. London: Cambridge University Press.

Tomasello, M. (2014). The ultra-social animal. European Journal of Social Psychology, 44(3), 187-194.

Turner, J. C., Hogg, M. A., Oakes, P. J., Reicher, S. D., & Wetherell, M. S. (1987). Rediscovering the social group: A self-categorization theory.

Cambridge, MA, US: Basil Blackwell.

Tyson, L. (1999). Critical theory today: A user-friendly guide. New York: Garland.

van Kuijk, I., Verkoeijen, P., Dijkstra, K., & Zwaan, R. A. (2018). The effect of reading a short passage of literary fiction on theory of mind: A replication of Kidd and Castano (2013). Collabra: Psychology, 4, 7. doi:10.1525/collabra.11

Walter, B. (1968). The work of art in the age of mechanical reproduction. Fontana.

Wiessner, P. W. (2014). Embers of society: Firelight talk among the Ju/'hoansi Bushmen. Proceedings of the National Academy of Sciences, 14027-14035.

Wrangham, R. (2009). Catching fire: how cooking made us human, Basic books, New York.

3. Emotional language processing in first (L1) versus second (L2) language

Francesco Foroni
Australian Catholic University , Australia

The ability to share, to communicate, and to understand emotions are fundamental for our social life. Humans communicate emotional information in many ways but, undoubtedly, one of the most important means is language. Language is a powerful elicitor of emotions and as such it can affect judgments (Johnson & Tversky, 1983) and, thus, have important implications also for face-to-face communication (Kawakami et al., 2007). Think about what happens when a friend is telling you that she has never been so "happy" or when you read about it in a text message or when we read in a book about someone who is very sad for the loss of someone. There is a large body of research on language processing looking at how we encode, store, and retrieve emotional verbal material (Kazanas et al., 2019). However, one interesting area of research investigates how we process emotional language in first language (L1) versus a second language (L2). Considering that we live and communicate more and more in multilingual and multicultural contexts, understanding possible differences in emotional language processing between L1 and L2 is paramount for the possible implications for emotion communication and cross-cultural communication (see Foroni, 2022).

The present contribution will look at the way emotional words and language are differently processed across L1 and L2 reviewing a selection of key behavioural, physiological, and brain imaging evidence. Finally, it will look at possible real-life implications of such differences.

Emotional language processing differences between L1 and L2

Behavioural evidence

In cognitive sciences, bilingual processing of emotional language/ words has been investigated extensively implementing a multitude of paradigms. When interested in the possible difference in the intensity of the emotion experience between L1 and L2, for instance, participants may be required first to evaluate the pleasantness of words presented in L1 and L2 and subsequently they may be tested in a free recall task for memory of the words previously evaluated. A standard result using this approach is that participants recall better emotional words compared to neutral words both in L1 and L2 (e.g., Ferré et al., 2010) suggesting that intensity of the emotional words in L1 and L2 may be similar. However, this is normally the case only when participants are early bilinguals (i.e., if the L2 is acquired early in life and the participants are very proficient in L2). Differences in intensity between L1 and L2, in fact, are reported when bilingual participants are not highly proficient in L2 and/or are living in an environment that is linguistically more focused on one of the languages (e.g., Altarriba & Bauer, 2004). This highlights how the differences in proficiency between L1 and L2 may be responsible for differences in emotional word processing and their emotional intensity.

Other paradigms that capitalize on interference effects have been implemented to test whether attention is directed to emotional stimuli differently in L1 and L2. Paradigms such as Rapid Serial Visual Presentation (e.g., Colbeck & Bowers, 2012), Affective Simon Task (e.g., Altarriba & Basnight-Brown, 2010), and Stroop Tasks (e.g., Sutton et al., 2007) provide generally a consistent picture and support the notion that emotional language and words capture our attention more readily in L1 compared to L2. Among others, the Stroop task is an interesting example. In the original Stroop task (Stroop, 1935), colour names (e.g., BLUE) are presented in an ink color that is congruent with word (i.e., blue ink) or incongruent (e.g., red ink) and the task of the participants is to name the ink color as quickly and as accurately as possible ignoring the word itself. When the word is presented in an incongruent ink color (word BLUE in red ink) performance is disrupted compared to when word and ink colour are congruent (word BLUE in blue ink). This original task has been successfully applied to the investigation of emotion words presenting words such as happy or sad in colored font to bilinguals. The emotional aspect of

words tends to capture attention and to interfere with the colour naming performance (Sutton et al., 2007).

Another way to investigate differences between L1 and L2 in emotion word representation has been to ask participants to rate words on dimensions like concreteness, imageability, or context-availability. For instance, Altarriba (2003) asked bilinguals to rate concrete (e.g., 'dog'), abstract (e.g., 'mind'), and emotion (e.g., 'happy') words in Spanish and English on those three dimensions. The results relative to the different word types on the three dimensions together suggest that emotion words in L1 are learned early within a complex learning environment and, so, they are encoded more richly and deeply compared to other types of words (i.e., concrete words). This evidence suggests that the learning context plays a larger role in terms of emotion word representation in L1 compared to L2. Notably, not only the learning history of L2 is important in explaining L1/L2 differences in emotion words processing. In fact, dominance of a language is not always referring to the first language learned (i.e., L1) for bilingual populations. Thus, the context moderates the L1/L2 difference beyond linguistic proficiency (Kazanas et al., 2019).

In summary, the general understanding from the behavioural literature is that emotional words are differently coded or characterized in our mental representations compared to other language and this has helped understanding the situations in which L1 (or L2) is the language in which emotions are more deeply coded for a bilingual speaker. Generally, emotions are tagged in L1 rendering L1 more emotional than L2. However, there is a caveat, namely, that emotional language needs to be learned, reinforced, and coded by the bilinguals at a time when L1 is the active or current language. If L2 words are newly coded and not deeply situated in memory, they tend not to have the same intensity and arousal of the more proficient L1 and this would be reflected in less interference in Stroop tasks or other tasks (cfr. Kazanas et al., 2019). Already from these results, one can speculate on how such L1/L2 differences could have different impact in applied real-life situations depending on the fact that the situation benefits from a stronger emotional load (e.g., clinical setting) or is impacted by such emotionality (e.g., decision-making).

Psychophysiological evidence

Physiological techniques (e.g., facial electromyography [EMG], skin conductance responses [SCRs]) have been paired with behavioural paradigms providing converging evidence and expanding our understanding

of the ways in which emotion processing in L1 compares with L2. This approach also allows to discriminate the nature of the emotional experiences. Generally speaking, EMG is generally considered a better technique to measure differences in valence, while SCRs is considered better suited for measuring arousal (e.g., Bradley & Lang, 2000; Lang et al., 1993; 1998).

Implementing facial EMG, Foroni and Semin (Foroni, 2015; Foroni & Semin, 2013) assessed the facial muscle activity of Dutch-English bilinguals while they were reading sentences that contained descriptions of emotional expressions in affirmative and negative form (e.g., I am smiling; I am not smiling). The authors were able to show that L1 processing engages the motor cortex more than L2. There were smaller muscle reactions in L2 compared to L1 and they concluded that processing emotional language in L1 likely relies on "simulations" of meaning and these muscle correlates are weaker in L2. Moreover, the authors reported that participants displayed qualitative similar facial muscle reactions in L1 and L2 while reading affirmative sentences. That is, even though there was a quantitative difference between L1 and L2 muscle activation with L1 showed larger muscle activation than L2, the patterns of muscle activations were qualitatively similar. On the other hand, while reading negative sentences in L1 and L2, bilinguals also display qualitative differences in their muscle reactions. In fact, when processing negative sentences in L2 bilinguals did not show any significant inhibition/relaxation of the relevant muscle as they instead showed for L1. If we consider muscle reactions while reading emotional language (i.e., activation/inhibition) as an index of somatic correlates of emotional language processing, then this latter result supports the claim that while emotional language processing in L1 relies on full simulations of the meaning described by the utterances, in L2 such simulations seem to occur only for part of the language forms (i.e., affirmative form; Foroni, 2015; Foroni & Semin 2013) providing evidence of an additional form of L1 advantage in emotional language processing.

Interestingly, weaker physiological reactions in L2 compared to L1 have also been reported for SCRs (e.g., Baumeister et al. 2017; Caldwell-Harris & Ayçiçeği-Dinn 2009; Harris et al., 2003). In a recent work Baumeister et al. (2017) examined EMG activity and SCRs of bilinguals while they were performing a categorization task (i.e., determining if a word is associated/not associated with emotions) and a recognition task. In line with the work reviewed in the behavioural evidence section, partic-

ipants categorized emotional words vs neutral words more efficiently in L1 compared to L2 and this effect could not be explained by differences in word fluency. This pattern was paralleled also by the results on a delayed recognition task. In line with these behavioural results, the authors also reported larger facial EMG responses and stronger SCRs for emotional words in L1 compared to L2 supporting the idea of an L1 advantage. Other literature implementing SCRs data also largely support an L1 advantage, with higher SCRs for emotional information presented in a bilingual's L1 (e.g., Caldwell-Harris & Ayçiçeği-Dinn, 2009; Harris et al., 2003; 2006).

Together, these physiological investigations provide evidence of a less extreme processing of emotional words in L2 with bilinguals experiencing reduced physiological responses compared to L1 emotion processing (Eilola & Havelka, 2010). However, this result is generally reported when L2 and L1 learning environments are not similar/well-matched (Kazanas et al., 2019) suggesting that part of the advantage is due to the differences in learning histories of L1 and L2.

Neuroscientific evidence

The growing implementation of neuroscientific techniques (e.g., Electroencephalogram [EEG] and functional Magnetic Resonance Imaging [fMRI]) has provided additional means to investigate emotional language/words processing across L1 and L2. On one hand, EEG data provide unique insights in the time course of cognitive processes across L1 and L2; on the other hand, fMRI investigations provide fine-grained information regarding localization and magnitude of the brain activity involved in emotional language processing in L1 versus L2.

In investigation focusing on event-related potentials (ERPs) using electroencephalogram (EEG) recordings (e.g., Conrad et al., 2011), two main findings are reported across L1 and L2. First, reading emotional words in L1 and L2 seems to induce a spontaneous activation of a word's emotional connotation characterised by the so-called early posterior negativity (EPN) in the left temporo-occipital electrode sites. This component is delayed by 50–100 ms in L2 compared to L1. Interestingly, the ERPs effects seem to be modulated also by the valence of the emotion (i.e., positive vs negative) and the language (German and Spanish in Conrad et al., 2011). Even though other investigations suggest that these differences in emotion activation do not seem to be modulated by difference in proficiency, frequency of use, or other relevant multilingual factors (Opitz & Degner, 2012), the results by Conrad et al. (2011) suggest the need to

further expand our understanding by systematically consider multiple languages as L1 and as L2 as there may be language-specific findings in these paradigms.

Secondly, later in the time course, L1 and L2 processing are distinguished by another ERP component called N400: i.e., a negative wave picking approximately around 400 ms post-stimulus that is detected generally across centroparietal electrode sites. N400 is often interpreted as a component reflecting conceptual/semantic incongruence/congruence in a sequence of stimuli (Kutas & Federmeier, 2011; Kutas & Hillyard, 1980a; 1980b). This semantic integration is usually observed in tasks like the Stroop task where participants must process competing information. Fan et al. (2016) tested late bilinguals implementing one of these tasks. Participants were presented with emotion words together with emotional faces creating congruent trials (e.g., word 'angry' presented with an angry face) and incongruent trials (e.g., the word 'angry' with and happy face). The authors reported that the interference between competing information was more readily apparent in L1 (i.e., larger N400), as more attention is given to emotional information in the conflict in L1.

These data, together with other research (e.g., Jończyk et al., 2016), suggest again a difference across L1 and L2; namely, late bilinguals experience their L2 in an incomplete, shallow manner. The results implementing ERPs methodology that taps the early time course of word processing are in line with the results described above implementing techniques that capture also slower, more effortful emotion processing like SCRs providing a consistent picture of L1/L2 differences in emotional words processing.

The implementation of fMRI technique has provided additional information and insights but the results are not as cohesive as those implementing behavioural and physiological techniques. Using fMRI technique, a portion of the literature does support an L1 advantage as evidenced by an increased brain activity while processing emotional information in L1 (e.g., Hernandez, 2009; Hsu et al., 2015). Emotional information processing is generally correlated with greater amygdala activation -- a region highly activated when processing emotion regardless of the positive or negative valence (Garavan et al., 2001; Hamann & Mao, 2002; Hamann et al., 2002) -- and interhemispheric communication (Jończyk, 2015). While amygdala shows increased activation for L1, interhemispheric communication is similarly engaged by emotional language irrespective of the language (L1 or L2). When comparing the activation across L1 and L2 it

seems clear that L1 may benefit from more widespread neural activity (e.g., Hernandez, 2009). Other studies report an L1 advantage in emotion processing also in other brain regions, with greater activity in the bilateral visual cortices, left precentral gyrus, and amygdala (Hsu et al., 2015). In general, this type of work tends to support the notion that a L1 may be more emotional compared to L2 even for early bilinguals that have a great deal of experience with both languages like was the case in the work by Hernandez (2009).

On the other hand, other studies showed no difference between L1 and L2 in the patterns of activation when processing emotional information (e.g., Yang et al., 2017). Few studies even report different patterns of activity altogether (e.g., Chen et al., 2015). The fMRI data reported by Chen et al. (2015), for instance, suggest qualitative differences in emotion processing between L1 and L2 rather than difference in degree of activation. Of the regions showing relevant activation for emotional words (e.g., left superior frontal gyrus, middle occipital gyrus, and left cerebellum), several patterns of activation differed across L1 and L2 processing. It is important to note that despite the fact that these qualitative differences are not necessarily in line with a simple L1 advantage explanation, they still suggest general differences in emotional language processing across L1 and L2.

Taken together, the brain imaging evidence largely supports an L1 advantage in emotion processing and are in line with the findings from psychophysiological studies measuring facial muscle activity and SCRs and behavioural research. However, one important conclusion from the evidence across paradigms is that L1 advantage may take different forms. In addition to the greater physiological reactivity reported with EMG and SCRs, L1 advantage appears in the form of earlier electrophysiological activity (ERPs) and more widespread neural activity (fMRI).

This imaging literature suffers from some limitations due to its recent development. Several of the reported results, for instance, are based on a small sample of late bilinguals whose language-learning experience differs across L1 and L2 (e.g., Caldwell-Harris & Ayçiçeği-Dinn, 2009; Caldwell- Harris et al., 2010; Harris et al., 2003; 2006). Despite the limitation, L1 advantage does persist across a wide variety of measures and experimental stimuli suggesting that it is an important phenomenon to further investigate. Thus, the obvious next step in this domain is to develop investigations with different populations of bilinguals and a wider array of emotional language material to further understand differences between

L1 and L2 in emotional language processing and possible boundary conditions (Kazanas et al., 2019).

Implications of differences between L1 and L2

As reviewed above, behavioural results as well as findings from psychophysiological and neuroimaging work largely indicate differences in the processing of emotional language in L1 and L2 and, generally, support an L1 advantage. Such differences and the L1 advantage can have significant implications for emotion communication in a range of domains considering our current multilingual and multicultural contexts. Initial evidence supports the conclusion that there are important real-life implications of L1/L2 difference. This evidence is related to clinical practice, forensic, decision-making, and marketing contexts (for a review, see also Caldwell-Harris, 2015).

In the field of counselling and clinical psychology, language is a primary means through which emotions are labelled, expressed, and discussed with clients (Altarriba et al., 1999). Moreover, it has been also suggested (De Zulueta, 2006) that the language we speak has an intrinsic link to our sense of identity, and using different languages is associated with experiencing changes in self-image (Pavlenko, 2006). Within clinical practice it has been known for some time that bilinguals sometimes seem to use L2 to distance themselves during the discussion of traumatic or troubling events (Altarriba, 2008; Pérez Foster, 1998; Pitta et al., 1978). Thus, in this area, the findings derived from behavioural, psychophysiological, and neuroimaging studies provide a strong base for this idea and can be directly utilized to improve mental health services for bilinguals. Santiago-Rivera and Altarriba (2002) reviewed evidence of the advantage of an emotionally distant L2 for the description of traumatic events. Thus, based on the available evidence a skilled therapist could assess the situation and lead the patient to switch between L1 and L2 in a way that best allows emotions to be accessed and addressed (Santiago-Rivera & Altarriba, 2002) suggesting L1/L2 switching as a possible emerging therapeutic technique.

Processing emotional language differently in L1 and L2 has also significant forensic implications. In the domain of morality, bilinguals consider a transgression as less harsh when the scenario is presented in L2 compared to when it is presented in L1 (Geipel et al., 2015). Looking at the domain of 'confrontation' Caldwell-Harris and Ayçiçeği-Dinn (2009)

showed that participants report more affective discomfort while lying in L1 compared to L2. This reduced emotionality in L2 could, then, promote suspects to engage in more lies and even be more likely to make false confessions when being interrogated in L2 (Kazanas et al., 2019).

Obviously, since L2 promotes greater cognitive and emotional distance, using L2 may provide an advantage in situations where lowering the emotionality level may be desirable. Keysar et al. (2012) have identified few decision-making biases that are indeed smaller in L2 because of the lower emotional involvement (see also Costa et al., 2014).

Finally, in the marketing area, Puntoni et al. (2009) reported that marketing slogans are perceived as less emotional in L2 than L1 suggesting a possible limitation of the effectiveness of an Ad in L2 if the Ad is based on strategically eliciting emotion reactions.

Together, these considerations suggest that the differences in emotional language processing between L1 and L2 seem to be a very promising and potentially important area for future investigations also because of the important real-life implications these differences may have in our current multilingual and multicultural societies.

References

Altarriba, J. (2003). Does cariño equal "liking"? A theoretical approach to conceptual nonequivalence between languages. *International Journal of Bilingualism*, *7*(3), 305-322.

Altarriba, J. (2008). Expression of emotion as mediated by context. *Bilingualism: Language and Cognition, 11*(2), 165-167.

Altarriba, J. & Basnight-Brown, D.M. (2010). The representation of emotion vs. emotion- laden words in English and Spanish in the affective Simon task. International Journal of Bilingualism 15: 310–328.

Altarriba, J., & Bauer, L. M. (2004). The distinctiveness of emotion concepts: A comparison between emotion, abstract, and concrete words. *American Journal of Psychology, 117,* 389-410.

Altarriba, J., Bauer, L. M., & Benvenuto, C. (1999). Concreteness, context availability, and imageability ratings and word associations for abstract, concrete, and emotion words. *Behavior Research Methods*, *31*(4), 578-602.

Baumeister, J. C., Foroni, F., Conrad, M., Rumiati, R. I., & Winkielman, P. (2017). Embodiment and emotional memory in first v. second language. *Frontiers in Psychology, 8*, 394.

Bradley, M.M., & Lang, P.J. (2000). Affective reactions to acoustic stimuli. *Psychophysiology* 37 (2): 204–215.

Caldwell-Harris, C. L. (2015). Emotionality differences between a native and foreign language: Implications for everyday life. *Current Directions in Psychological Science, 24*(3), 214-219.

Caldwell-Harris, C. L., & Ayçiçeği-Dinn, A. (2009). Emotion and lying in a non-native language. *International Journal of Psychophysiology, 71*(3), 193-204.

Caldwell-Harris, C. L., Tong, J., Lung, W., & Poo, S. (2010). Physiological reactivity to emotional phrases in Mandarin-English bilinguals. *International Journal of Bilingualism, 15*(3), 329-352.

Chen, P., Lin, J., Chen, B., Lu, C., & Guo, T. (2015). Processing emotional words in two languages with one brain: ERP and fMRI evidence from Chinese–English bilinguals. *Cortex, 71*, 34-48.

Colbeck, K. L., & Bowers, J. S. (2012). Blinded by taboo words in L1 but not L2. *Emotion, 12*(2), 217-222.

Conrad, M., Recio, G., & Jacobs, A. M. (2011). The time course of emotion effects in first and second language processing: A cross-cultural ERP study with German–Spanish bilinguals. *Frontiers in Psychology, 2*, 351.

Costa, A., Foucart, A., Arnon, I., Aparici, M., & Apesteguia, J. (2014). "Piensa" twice: On the foreign language effect in decision making. *Cognition, 130(2)*, 236-254.

De Zulueta, F. (2006). *From pain to violence: The traumatic roots of destructiveness*. London, UK: John Wiley & Sons.

Eilola, T. M. & Havelka, J. (2010). Behavioural and physiological responses to the emotional and taboo Stroop tasks in native and non-native speakers of English. *International Journal of Bilingualism, 15*(3), 353-369.

Fan, L., Xu, Q., Wang, X., Zhang, F., Yang, Y., & Liu, X. (2016). Neural correlates of task-irrelevant first and second language emotion words: Evidence from the emotional face–word Stroop task. *Frontiers in Psychology, 7*, 1672.

Ferré, P., García, T., Fraga, I., Sánchez-Casas, R., & Molero, M. (2010). Memory for emotional words in bilinguals: Do words have the same emotional intensity in the first and in the second language? *Cognition and Emotion, 24*(5), 760-785.

Foroni, F. (2015). Do we embody second language? Evidence for 'partial' simulation during processing of a second language. *Brain and Cognition, 99*, 8-16.

Foroni, F. (2022). Multimodality, facial expression, and emotional

language. In G.L. Schiewer, J. Altarriba, & B.C. Ng (Eds.). *Handbook on Language and Emotion* (pp. 364-384). De Gruyter Mouton, Berlin, Germany.

Foroni, F., & Semin, G. R. (2013). Comprehension of action negation involves inhibitory simulation. *Frontiers in Human Neuroscience, 7*, 209.

Garavan, H., Pendergrass, J. C., Ross, T. J., Stein, E. A., & Risinger, R. C. (2001). Amygdala response to both positively and negatively valenced stimuli. *NeuroReport, 12*, 2779–2783.

Geipel, J., Hadjichristidis, C., & Surian, L. (2015). How foreign language shapes moral judgment. *Journal of Experimental Social Psychology, 59*, 8-17.

Hamann, S. B., Ely, T. D., Hoffman, J. M., & Kilts, C. D. (2002). Ecstasy and agony: Activation of the human amygdala in positive and negative emotion. *Psychological Science, 13*, 135–141.

Hamann, S., & Mao, H. (2002). Positive and negative emotional verbal stimuli elicit activity in the left amygdala. *NeuroReport, 13*, 15–19.

Harris, C. L., Ayçiçeği, A., & Gleason, J. B. (2003). Taboo words and reprimands elicit greater autonomic reactivity in a first language than in a second language. *Applied Psycholinguistics, 24*, 561-579.

Harris, C. L., Gleason, J. B., & Ayçiçeği, A. (2006). When is a first language more emotional? Psychophysiological evidence from bilingual speakers. In A. Pavlenko (Ed.). *Bilingual minds: Emotional experience, expression, and representation* (pp. 257-283). Clevedon, UK: Multilingual Matters.

Hernandez, A. E. (2009). Language switching in the bilingual brain: What's next? *Brain & Language, 109*(2), 133-140.

Hsu, C-T., Jacobs, A. M., & Conrad, M. (2015). Can Harry Potter still put a spell on us in a second language? An fMRI study on reading emotion-laden literature in late bilinguals. *Cortex, 63*, 282-295.

Johnson, E. J., & Tversky, A. (1983). Affect, generalization, and the perception of risk. *Journal of personality and social psychology, 45*(1), 20.

Jończyk, R. (2015). Hemispheric asymmetry of emotion words in a non-native mind: A divided visual field study. *Laterality, 20*(3), 326-347.

Jończyk, R., Boutonnet, B., Musial, K., Hoemann, K., & Thierry, G. (2016). The bilingual brain turns a blind eye to negative statements in the second language. *Cognitive, Affective, and Behavioral Neuroscience, 16*, 527-540.

Kawakami, K., Phills, C. E., Steele, J. R., & Dovidio, J. F. (2007). (Close)

distance makes the heart grow fonder: Improving implicit racial attitudes and interracial interactions through approach behaviors. *Journal of personality and social psychology*, *92*(6), 957.

Kazanas, S. A., McLean, J. S., & Altarriba, J. (2019). Emotion and emotion concepts: Processing and use in monolingual and bilingual speakers. *The handbook of the neuroscience of multilingualism*, 313-334.

Keysar, B., Hayakawa, S. L., & An, S. G. (2012). The foreign-language effect: Thinking in a foreign tongue reduces decision biases. *Psychological science*, *23*(6), 661-668.

Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: finding meaning in the N400 component of the event-related brain potential (ERP). *Annual Review of Psychology, 62*, 621–647.

Kutas, M., & Hillyard, S. A. (1980a). Event-related brain potentials to semantically in- appropriate and surprisingly large words. *Biological Psychology, 11(2)*, 99–116.

Kutas, M., & Hillyard, S. A. (1980b). Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science, 207(4427)*, 203–205.

Lang, P.J., Bradley, M.M., & Cuthbert, B.N. (1998). Emotion, motivation, and anxiety: brain mechanisms and psychophysiology. *Biological Psychiatry* 44 (12): 1248–1263.

Lang, P.J., Greenwald, M.K., Bradley, M.M., & Hamm, A.O. (1993). Looking at pictures: affective, facial, visceral, and behavior reactions. *Psychophysiology* 30: 261–273.

Opitz, B., & Degner, J. (2012). Emotionality in a second language: It's a matter of time. *Neuropsychologia, 50*(8), 1961-1967.

Pavlenko, A. (2006). Bilingual selves. In A. Pavlenko (Ed.), *Bilingual minds: Emotional experience, expression and representation* (pp. 1-33). Clevedon: Multilingual Matters.

Pérez-Foster, R. M. (1998). *The power of language in the clinical process*: *Assessing and treating the bilingual person.* Northvale, NJ: Jason Aronson.

Pitta, P., Marcos, L. R., & Alpert, M. (1978). Language switching as a treatment strategy with bilingual patients. *The American Journal of Psychoanalysis, 38,* 255-258.

Puntoni, S., De Langhe, B., & Van Osselaer, S. M. (2009). Bilingualism and the emotional intensity of advertising language. *Journal of consumer research*, *35*(6), 1012-1025.

Santiago-Rivera, A. L., & Altarriba, J. (2002). The role of language in therapy with the Spanish-English bilingual client. *Professional Psychology: Research and Practice, 33*(1), 30-38.

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology, 18*(6), 643-662.

Sutton, T.M., Altarriba, J., Gianico, J.L., & Basnight-Brown, D.M. (2007). The automatic access of emotion: emotional Stroop effects in Spanish–English bilingual speakers. *Cognition and Emotion* 21 (5): 1077–1090.

Yang, Y., Wang, J., Bailer, C., Cherkassky, V., & Just, M. A. (2017). Commonality of neural representations of sentences across languages: Predicting brain activation during Portuguese sentence comprehension using an English-based model of brain function. *NeuroImage, 146*, 658-666.

4. Cultural Influences on Memory

Namrata Goyal
ESADE Business School, Spain

Culture and memory: attention, filtering, and categorization.

In our complex societies, we are exposed to a considerable amount of social information on a daily basis. Humans, however, are not passive in information uptake. We take an active role in selecting, filtering, and organizing information. Sometimes these processes are deliberate and effortful; we chose what to attend to. However, often we select, filter, and organize information spontaneously, placing this information into "memory buckets" that psychologists call mentally constructed social categories (Taylor et al., 1978; Billig & Tajfel, 1973; Tajfel & Wilkes, 1963).

Cognitive psychologists argue that information within such mentally constructed social categories influences memory more than information outside of these categories (Billig & Tajfel, 1973; Tajfel & Wilkes, 1963). For example, have you ever accidentally called your lover by the name of your ex-lover? Interchanged the names of your two children while speaking to them? Or mistakenly sent a sensitive email to your ex-boss instead of your current boss? We often joke about such errors calling them "Freudian slips" that we believe reveal our innermost feelings. However, cognitive psychologists believe that these errors occur because people are more prone to confuse information within their own mentally constructed social categories, leading to within-category errors, rather than making mistakes between different categories. To elaborate, within-category errors might involve mistakenly recalling a conversation with one friend instead of another who shares similar characteristics and interests. In this

case, both individuals belong to the same social category (friends with similar traits), making it more likely for errors to occur within this category. On the other hand, between-category errors would involve confusing information from different mental categories, which is considered less common based on cognitive psychological theories. Understanding these distinctions sheds light on how our cognitive organization of social information influences the accuracy of our memory recall.

While psychologists have been investigating individual differences that may influence memory processes for centuries, in the last few decades, cognitive psychologists have begun to systematically explore the way that culture may affect memory via attention, filtering, and categorization. Dr. Anne Maass has been a leading figure in this field. She has made groundbreaking discoveries about the influence of language and culture in memory recall of visual and linguistic information. In this chapter, I will review existing research that has assessed cultural influences on memory.

Culture may influence memory in different ways. Culture may channel attention thereby influencing memory via the weight we assign to information (Chua, Boland, & Nisbett, 2005; Imada & Yus- sen, 2012; Masuda & Nisbett, 2001; Nisbett & Masuda, 2003). For example, in a series of experiments Masuda and Nisbett (2001) exposed participants to an underwater scene with different objects finding that East Asians attended more to the context, but that Americans focused more on focal objects. These differences in attention resulted in different memory of identical information. Americans were more likely than Asians to remember individual objects in the scenes. These researchers hence concluded that whereas East Asians tend to remember information in a more holistic manner, Americans are prone to analytical processing of information. The observed differences in attention and memory between East Asians and Americans, as highlighted by Masuda and Nisbett (2001), are often attributed to the cultural dimensions of individualism and collectivism. Collectivist cultures prioritize group coherence and relationships, leading individuals to focus on the overall context of a situation. As a result, East Asians tend to adopt a holistic approach to processing information, considering the interconnectedness of elements within a scene. Conversely, American culture tends to align with individualistic values, emphasizing personal autonomy and achievement. This cultural orientation may contribute to analytical processing, where individuals focus more on distinct and salient objects rather than the broader context. In the study's un-

derwater scene experiments, Americans exhibited a higher tendency to remember individual objects, reflecting their analytical processing style. Therefore, the observed cultural differences in attention and memory can be linked to the broader cultural frameworks of individualism and collectivism, shedding light on how cultural values shape cognitive processes and information processing strategies.

Culture may also influence memory by influencing the way we filter information (Gutchess, Welsh, Bodurogˆlu, & Park, 2006; Lewis, Goto, & Kong, 2008; Maass, Karasawa, Politi, & Suga, 2006; Morris & Mok, 2011; Na & Kitayama, 2011; Uleman, Adil Saribay, & Gonzalez, 2008; Wagar & Cohen, 2003, Bettinsoli et al., 2019). For example, Maass et al (2006) asked Italian and Japanese participants to read a recommendation letter, ostensibly written by a previous employer. The recommendation letter contained eight critical pieces of information. Participants received half of these in the form of adjectives (e.g., dominant) and half in the form of verbs (e.g., expects to be obeyed). Participants were then asked to recall information from this letter. Maass et al (2006) found that Westerns (e.g., Italians) were more likely than Japanese to convert behavioral information (e.g., expects to be obeyed) into trait representations (e.g., dominant) when retrieving information about people, even when both cultural groups agreed that adjectives were more informative than verbs. These results build upon earlier cultural psychology research by Miller (1984) and Morris (1992, 1994), who explored cognitive biases in person perception. Notably, these studies revealed cross-cultural differences in the fundamental attribution error, with European American participants more inclined toward person attribution biases, inferring behavior as reflective of personality, while Chinese and Indian participants tended to provide contextual explanations, reflecting situational attribution biases (Choi & Nisbett, 1998; Norenzayan et al., 2002; Knowles, 2001; Benet-Martinez et al., 2003).

Research has also shown that culture can influence memory both via deliberate and spontaneous categorization (Boroditsky, Schmidt, & Phillips, 2003; Gutchess & Bodurogˆlu, 2018; Ji, Zhang, & Nisbett, 2004; Karasawa, Maass, Rakic´, & Kato, 2014; Schwartz, Bodurogˆlu, & Gutchess, 2014; Unsworth, Sears, & Pexman, 2005). Pioneering work by Chui (1972) was likely the first to document cultural differences in categorization strategies. Chiu found differences in categorization between Chinese and American children using different objects. In a series of studies, participants were presented with a set of pictures (e.g., tire, car and a bus) and

then asked to group the two pictures they felt best belonged together. Chui found that while American children more often grouped objects based on their similarity (e.g., a car and a bus), Chinese children were more likely to categorize based on the relationships between objects (e.g., tire and a car). These results imply that whereas Chinese may recall information based on relational similarities, Americans may be more inclined to recall information based on similarity. These cultural differences in memory categorization strategies may arise from variations in cultural norms, values, and cognitive frameworks that shape how individuals perceive and organize information. Cultural norms can influence the emphasis placed on relational versus similarity-based processing, impacting memory categorization strategies. Additionally, societal values regarding individualism and collectivism may play a role, with cultures emphasizing interconnectedness and relationships (as seen in Chiu's study) influencing cognitive processes differently than those prioritizing independence and individual attributes. Moreover, the linguistic and communicative patterns within a culture can contribute to divergent memory encoding strategies, ultimately affecting the way people recall and categorize information.

Karasawa et al (2014) however took this work one step further in documenting not just cultural differences in deliberate categorization, but also spontaneous categorization of information. Karasawa et al (2014) assessed differences between Japanese and Italians participants hypothesizing that perceived age differences among people would trigger spontaneous categorization of information for Japanese participants only, whereby people would mentally construct categories based on age. They tested their hypothesis using the "Who said what?" paradigm originally contrived by Taylor, Fiske, Etcoff & Ruderman (1978). The "Who said what?" paradigm, is an experimental method used to study social perception and attribution. In this paradigm, participants are presented with statements expressing various viewpoints, and their task is to attribute each statement to a specific group of people. The design aims to investigate how individuals categorize and attribute statements to social groups. Using this paradigm, Karsawa et al (2014) presented 19-22 year-olds Japanese and Italians participants with 6 profiles of individuals that varied based on age (younger age group "16, 17, 18 years" vs older age group "24, 25, 26 years"). Participants were presented with information about these profiles and then asked to complete a memory task at the end of the study. Karasawa et al (2014) found that only Japanese (but not Italians)

made "within-category" memory errors. That is, Japanese participants were more likely than Italian participants to make memory errors within the young prolife age group or the old profile age group (e.g., misrecognizing the 16-year-old target as the 18-year-old because they all belong to the younger-than-the-self" group), than make memory errors between age categories (e.g, the confusion between the 18-and 24-year-old).

In follow-up work that Anne Maass and I have done together, we found that culture may influence spontaneous categorization, even independent of language. We proposed that norms act in subtle, unconscious ways to influence our memory; norms guide us to spontaneously organize information on the basis of whether the information fits with or violates an activated norm, resulting in altered memory. In order to show this effect, we took advantage of existing cultural differences in normative behaviors; we deliberately sampled Indians and Americans as these groups have been shown to differ in the norms they utilize in various domains. We then identified two domains: reciprocity and purity. Americans but not Indians, normatively reciprocate help received via gifts (Miller et al 2014; Goyal & Miller 2017), and Indians but not Americans, normatively view shoes as spiritually contaminating objects and thus do not allow shoes to come into contact with sacred objects (Shweder, Much, Mahapatra, Park, 1997). We hypothesized that Indians and Americans would have different memory for the same information as their culturally variable norms spontaneously guide these two groups to sort information differently. We found support for this hypothesis in four studies. In some experiments, we activated purity norms regarding shoes among Indians (but not Americans). We assessed whether identical shoe advertisements would result in memory differences between Indians and Americans. All participants were presented with three advertisements in which shoes were displayed on sacred objects (i.e., books, a dinner plate, and a cow) and three in which shoes were displayed on neutral objects (a clock, tools, and a telephone). We found that Indians (but not Americans) grouped together in their memory the shoes that appeared on sacred objects, as they interpreted these images as norm violations, and also grouped together shoes that appeared on neutral objects as they interpreted these as non-violations. Thus, when asked to recall the advertisements, Indians (but not Americans) made systematic memory errors: they confused details of the advertisements within each group but not between the groups. Since Americans did not interpret any of the advertisements as norm-relevant, they did not make systematic memory errors.

Together our studies found that cultural norms operate by unconsciously guiding individuals to sort through information based on whether or not information violates activated norms. Thus, this work demonstrates cultural norms act automatically to influence how we organize information resulting in altered memory.

Implications for research on culture and memory

The research on culture and memory presents a rich tapestry of insights with profound implications for our understanding of human cognition and social interactions. The fundamental idea that individuals actively engage in the selection, filtering, and organization of information, both consciously and spontaneously, has far-reaching consequences for various aspects of daily life. These insights, rooted in the spontaneous interplay of cultural norms and memory, shed light on the intricate ways in which cultural contexts shape cognitive processes, offering practical implications for diverse fields ranging from interpersonal communication to conflict management. For example, understanding the impact of cultural norms on memory can significantly enhance interpersonal communication. Individuals from different cultures may prioritize distinct aspects of information, leading to potential misunderstandings. Recognizing these cultural variations enables people to navigate cross-cultural interactions more effectively, fostering mutual understanding and cooperation. In the realm of conflict mitigation, understanding that different cultures may emphasize distinct aspects of information can help individuals engaged in conflicts recognize that diverse cognitive processes influence perceptions and interpretations. By acknowledging and respecting varied cultural approaches to memory, individuals can approach conflicts with greater sensitivity and openness. This awareness can pave the way for constructive dialogue, where conflicting parties appreciate the potential influence of cultural factors on their perspectives. Moreover, recognizing the role of cultural norms in shaping memory can aid in identifying and addressing potential triggers for misunderstandings, ultimately contributing to more effective conflict resolution strategies. Overall, the exploration of culture and memory opens avenues for enhancing cultural awareness, understanding, and adaptability in various spheres of human interaction.

Concluding thoughts

In this article, I have reviewed some exciting work on how culture may influence memory by influencing the uptake of information via attention, filtering, and categorization, in the present. However, a large body of evidence has found systematic cultural differences even for memories of the past, i.e., autobiographical memory (for review see Wang 2021). Culture may thus contribute not only to what, when, and how people remember information but also to whether people judge remembering to be important at all.

References

Billig, M., & Tajfel, H. (1973). Social categorization and similarity in intergroup behavior. European Journal of Social Psychology, 3, 27–52. http://dx.doi.org/10.1002/ejsp.2420030103

Boroditsky, L., Schmidt, L. A., & Phillips, W. (2003). Sex, syntax, and semantics. In D. Genter & S. Goldin-Meadow (Eds.), Language in mind: Advances in the study of language and thought (pp. 61–79). Cambridge, MA: The MIT Press.

Chua, H. F., Boland, J. E., & Nisbett, R. E. (2005). Cultural variation in eye movements during scene perception. Proceedings of the National Academy of Sciences of the United States of America, 102, 12629–12633. http://dx.doi.org/10.1073/pnas.0506162102

Chiu, L. (1972). A cross-cultural comparison of cognitive styles in Chinese an American children. International Journal of Psychology, 7, 235-242.

Goyal, N., & Miller, J. G. (2018). The importance of timing in reciprocity: An investigation of reciprocity norms among Indians and Americans. Journal of Cross-Cultural Psychology, 49, 381–403. http://dx.doi.org/10.1177/0022022117746239

Goyal, N., Adams, M., Cyr, T. G., Maass, A., & Miller, J. G. (2020). Norm-based spontaneous categorization: Cultural norms shape meaning and memory. Journal of personality and social psychology, 118(3), 436.

Gutchess, A. H., Welsh, R. C., Boduroĝlu, A., & Park, D. C. (2006). Cultural differences in neural function associated with object processing. Cognitive, Affective & Behavioral Neuroscience, 6, 102–109. http://dx.doi.org/10.3758/CABN.6.2.102

Gutchess, A. H., Welsh, R. C., Boduroĝlu, A., & Park, D. C. (2006). Cultural differences in neural function associated with object processing.

Cognitive, Affective & Behavioral Neuroscience, 6, 102–109. http://dx.doi.org/10.3758/CABN.6.2.102

Imada, T., & Yussen, S. R. (2012). Reproduction of cultural values: A cross-cultural examination of stories people create and transmit. Personality and Social Psychology Bulletin, 38, 114–128. http://dx.doi.org/10 .1177/0146167211421938

Ji, L. J., Zhang, Z., & Nisbett, R. E. (2004). Is it culture or is it language? Examination f language effects in cross-cultural research on categorization. Journal of Personality and Social Psychology, 87, 57–65. http:// dx.doi.org/10.1037/0022-3514.87.1.57

Karasawa, M., Maass, A., Rakić, T., & Kato, A. (2014). The emergent nature of culturally meaningful categorization and language use: A Japanese–Italian comparison of age categories. Journal of Cross-Cultural Psychology, 45, 43-451. http://dx.doi.org/10.1177/ 0022022113509882

Masuda, T., & Nisbett, R. E. (2001). Attending holistically versus analytically: Comparing the context sensitivity of Japanese and Americans. Journal of Personality and Social Psychology, 81, 922–934. http://dx.doi.org/10.1037/0022 3514.81.5.922

Miller, J. G., Bland, C., Källberg-Shroff, M., Tseng, C. Y., Montes-George, J., Ryan, K., Chakravarthy, S. (2014). Culture and the role of exchange vs. communal norms in friendship. Journal of Experimental Social Psychology, 53, 79–93. http://dx.doi.org/10.1016/j.jesp.2014.02.006

Nisbett, R. E., & Masuda, T. (2003). Culture and point of view.

Proceedings of the National Academy of Sciences of the United States of America, 100, 11163–11170. http://dx.doi.org/10.1073/pnas .1934527100

Schwartz, A. J., Bodurogˆlu, A., & Gutchess, A. H. (2014). Cross-cultural differences in categorical memory errors. Cognitive Science, 38, 997–1007. http://dx.doi.org/10.1111/cogs.12109

Tajfel, H., & Wilkes, A. L. (1963). Classification and quantitative judgment. British Journal of Psychology, 54, 101–114. http://dx.doi.org/10 .1111/j.2044 8295.1963.tb00865.x

Taylor, S. E., Fiske, S. T., Etcoff, N. L., & Ruderman, A. J. (1978). Categorical and contextual bases of person memory and stereotyping. Journal of Personality and Social Psychology, 36, 778–793. http://dx .doi.org/10.1037/0022-3514.36.7.778

Unsworth, S. J., Sears, C. R., & Pexman, P. M. (2005). Cultural influences on categorization processes. Journal of Cross-Cultural Psychology, 36, 662-688. http://dx.doi.org/10.1177/0022022105280509

5. You Broke It or It Broke? A Cross-Linguistic Analysis of Verb Transitivity in Causal Explanations

Minoru Karasawa[1] and Yuko Yoshinari[2]
[1]Nagoya University, Japan
[2]Gifu University, Japan

In January 2003, an international symposium of psychology and language was held at Kobe University in Japan. The invited speakers included Anne Maass, Gün Semin, and Shin'ichiro Okamoto from psychology as well as Professors Sachiko Ide and Yoshihiko Ikegami from linguistics. Inspired by the symposium, a graduate student in the audience made the decision that Kobe was the place to which she should transfer and pursue her goal of incorporating psychological approaches into her research in linguistics. Years later, she is the second author of this chapter. The symposium also led the co-organizers, the late Yoshihiro Nishimitsu and the first author, to launch a joint graduate seminar for students majoring in psychology and in linguistics.

One product of this interdisciplinary seminar was a research project on the use of transitive versus intransitive verbs and the perception of responsibility. Linguists characterize verb transitivity by its association with agentivity, typically defined by factors such as the causality, volitionality, and specificity in influence of a behavioral event (e.g., Hopper & Tompson, 1980). These grammatical "rules," however, are typically based on the syntactic structure found in Western languages. It should not be surprising, therefore, to find utterances with similar pragmatic aims take different syntactic forms or different markers between different languages (e.g., Jacobsen, 1991). As we will discuss below, analyses of verb transi-

tivity provide an interesting empirical tool for us to examine differences and similarities in how speakers of different languages view social events.

Agent- vs. Context-Oriented Languages

A predominant view in current theories of cultural psychology pertains to cultural variations in the allocation of attention to the focal object as compared with that to its surrounding environment (Miyamoto, 2013). The tendency to pay greater attention to the focal target in a "field-independent" manner is called analytic cognition which is typically prevalent in North America and West Europe. The other type called holistic cognition takes into consideration a greater deal of the background context and is more often found in areas like East Asia. Evidence also shows that the difference between analytic versus holistic cognition reflects in differential emphasis on a causal agent as opposed to the situation as a whole (Nisbett, Peng, Choi, & Norenzayan, 2001).

The above distinction in different cultural styles of cognition resonates with a contrast we find in linguistic studies. Specifically, Ikegami (1991) maintained that Indo-European languages (with English as a prominent case) can be characterized as "DO-languages," whereas other languages (Japanese in particular) as "BECOME-languages." According to this distinction, a primary feature of DO-languages lies in their emphasis on the agentivity. This typically reflects in the relatively explicit use of agentic markers such as the subject (S) and object (O) in the described event, and the syntactic rules predicating the S-O relationship by a verb (V). In contrast, he argues, BECOME-languages tend to be structured to represent the thematic context of how the event took place and what is the consequential state of the event. Consistent with this characterization is the frequent use of *wa*, a topical particle standing for "as for," in the Japanese language. Another well-known characteristic of Japanese is that it allows sentence subjects to be missed so commonly, leaving out the agentivity to be inferred from the context. Certainly, we find this phenomenon of "pronoun drop" in Western languages as well to varying degrees, but it usually accompanies a verb conjugation which explicitly marks the S. However, east-Asian languages including Japanese and Korean, maintain the same form of V, and therefore, the dropped S needs to be inferred. (For a cross-linguistic analysis on psychological correlates of pronoun drop, see Kashima & Kashima, 1998).

Linguists generally concur that agentivity is more clearly marked in English than in Japanese. For instance, Hinds (1986) proposed a distinction between English and Japanese as Person-Focused versus Situation-Focused, respectively, which overlaps with Ikegami's characterization to a great extent. Taken together, linguistic analyses suggest that narratives in English and other Indo-European languages tend to express who did what (to whom), whereas other languages (Japanese, in particular) tend to describe what happened and how. This difference led Ikegami (1991) to conclude that the agentivity in Japanese is relatively "blurred or suppressed" (p. 309).

The emphasis on agentivity marked in English also leads to a peculiar linguistic characteristic, namely, demarcating a person. The following example in English illustrates this tendency of personification.

> ENG: I don't understand you.
> JPN: Anatano iu koto ga wakaranai.
>   youPOSS saying thing OBJ understand:NEG[1]
>   "(I) don't understand what you say."

The Japanese counterpart is unable to lack a precise referent of the mental state (i.e., not understanding) and would hence use "koto" to nominalize "what you say." Note also that the sentence is void of what is semantically equivalent to the subject in the English expression because of the convention of pronoun drop. In other words, it appears more important in the English utterance to state who understands whom whereas the content of the mental state of understanding plays the central role in Japanese.

Admittedly, not every event is caused by a human. When the cause is a non-human entity, the relatively strong tendency in English to search the causal agent allows even an inanimate object to be the sentence subject, as seen in the example below. In contrast, the Japanese language tends to avoid using an inanimate subject and, as a consequence, typically takes an intransitive or a passive form.

> ENG: The typhoon killed 36 people.
> JPN: Taifu de 36-mei ga nakunatta.
>   typhoon by 36-people SBJ die:PST
>   "Thirty-six people died by typhoon."

---

1 Abbreviations used to label linguistic terms in Japanese language examples: POSS possessive, NEG negative, OBJ Object, PROG progressive, PST past, SBJ Subject, TOP topic

In addition to causality, ownership is another important aspect of agentivity. DO-languages allow a relatively simple and systematic representation of ownership. It is worth noting that ownership is associated with the owner's property right accompanied by responsibility, both of which coincidentally are major characteristics of individualist values. (We will discuss the perception of responsibility in relation to the linguistic agentivity in the next section.) On the other hand, the syntactic form of "X has Y" does not always translate literally into a BECOME-language (Hinds, 1986).

> ENG: I have a fever today.
> JPN: Kyo wa netsu ga   aru.
>   today TOP fever SBJ exist
>   "There is a fever today."

Here too, the focus of the English narrative is the person, with the consequence of their health condition also implied (e.g., difficulty in working). The Japanese version may carry the same message, but the linguistic structure is noticeably different. The focal point is the existence of the health condition of the unmarked protagonist, whom we need to infer from the context.

Likewise, in the following pair, both narratives describe the same situation with certain implications for the consequence (e.g., embarrassment, the cost for cleaning, and so forth), but only the English version visibly marks the ownership with the S-V-O structure.

> ENG: I have a stain on the coat.
> JPN: Kooto ni shimi ga   tsuiteiru.
>   coat on   stain SBJ attach:PROG
>   "There is a stain on the coat."

To summarize, as Ikegami (1991) pointed out, English exhibits the tendency of DO-languages, expressing agentivity more explicitly in its grammatical features, such as the S-V-O structure and the use of transitive verbs, marking the causal agency and specifying ownership. In contrast, Japanese as a BECOME-language inclines toward the topical representation of a state or the intransitive aspect of an event along with a certain context. The difference is typically reflected in the use of transitive versus intransitive verbs in describing the event. Hence, we shall next attempt a thorough analysis of this linguistic variable and its psychological implications.

Verb Transitivity and the Perception of Responsibility

Both psychologists (Heider, 1958; Malle, Guglielmo, & Monroe, 2014; Weiner, 1986) and linguists (DeLancey, 1985; Ikegami, 1982; Nishimura, 1997) have pointed out that an important component of agentivity is the causality in the described event. Because of the syntactic rule requiring S, V, and O, it is easier to represent the causality in DO-languages. Furthermore, the word order can facilitate the causal inference, with the word that appears first, rather than the one appearing in a later position, assigned a greater causal role (Bettinsoli, Maass. Kashima, & Suitner, 2015). Studies have also revealed that the spatial position in a written form (i.e., either writing from left to right or vice versa) representing the word order influences the agent-patient inferences (i.e., the "Spatial Agency Bias"; Suitner & Maass, 2016).

A visible marker of agentivity is the transitivity of the predicates in utterances. From a psychological perspective, the transitivity in describing a behavioral event invites various interpretations of the event as to causality and intentionality. Particularly relevant to this point are the studies conducted by Fausey and her colleagues (Fausey & Boroditsky, 2011; Fausey, Long, Inamori, & Boroditsky, 2010). These researchers presented short video clips to participants and asked them to generate open-ended descriptions of what happened. Results showed that even when the depicted events were visibly "accidental" (e.g., turning around and stepping on a can on the floor), English speakers more frequently used transitive verbs than did Japanese and Spanish speakers. Furthermore, encoding the accidental events in transitive verbs resulted in better memory of the episodes, particularly among English speakers. Presumably, transitive verbs were helpful, particularly for those English speakers, who are assumed to be more agentivity-prone, to identify the agentic source of a behavioral event and retain the interpretation.

It should be noted that the studies by Fausey and colleagues pertained to a context where verbal expressions and inferences were made from the viewpoint of a third-person observer who had little involvement with the events. However, using transitive verbs in a different context may serve a different pragmatic function. One possibility is when the utterance involves interpersonal dependency.

Drawing on the assumed difference between English and Japanese speakers in their linguistic emphases, with the former more on identifying the causality while the latter on searching the appropriateness of interpersonal utterances in the given context (Hinds, 1986), one may expect

that the Japanese might prioritize politeness over the literal accuracy in defining an identifiable fact such as causality. An episode often cited in textbooks of Japanese as a second language resonates with this view: An American student borrowed an appliance from her landlord, but it went out of order while she was using it. When she returned it, the student said, "This appliance broke down," which upset the landlord. Later, her Japanese instructor told her that she should have said, "I broke the appliance," to acknowledge in whose possession the appliance was originally. Very likely, the landlord then would have said, "Don't worry, it was old anyway" (Mizutani, 1987). This anecdote exemplifies the possibility that accentuated transitivity may serve a goal of politeness, sounding more "natural" under certain circumstances in certain linguistic cultures.

The above discussion raises an intriguing psychological question concerning the extent to which the judgment of responsibility is related to the choice of transitive verbs along with its implied agentivity. The aforementioned anecdote implies that the relationship may vary across different linguistic cultures. For the American student, it was probably logical to use an intransitive expression in order to establish that the malfunction was not her fault. In contrast, the Japanese landlord likely interpreted the comment as an attempt to evade responsibility because to him, linguistically marking the causality was less important to their relationship.

In addition to the internal versus external locus of causality, another potential determinant of perceived responsibility is the controllability of the cause, according to the attribution theory of responsibility (Weiner, 1986). In his linguistic analysis, Ikegami (1982) also pointed out that the "preventability," conceptually similar to controllability, should be a primary factor of responsibility implied by the use of transitive verbs. In the section below, we summarize our own findings concerning the relationship between the choice of descriptive verbs and causal inference along with judgment of responsibility, with special attention to peculiar effects observed among Japanese speakers.

An Empirical Study of Verb Transitivity in Japanese

In one of our earlier studies, participants (Japanese undergraduates) read hypothetical scenarios taking the viewpoint of the main protagonists. Each scenario depicted a damage that incurred to the protagonist's friend. We varied the causal information included in each scenario to

manipulate the *locus* (internal vs. external to the protagonist) and *controllability* of the cause of the damage as within-participant variables.

As summarized in Table 1, we prepared two sets of scenarios. The first "coffee scenario" started with this background information: "Suppose you were reading a book that you borrowed from a friend," followed by the experimental manipulation. The scenario in the Internal/Controllable condition read in an intransitive form: "While reading, your hand *hit on* the mug, and then... [i.e., *te ga atari*]." The Internal/Uncontrollable scenario added, "You suddenly became dizzy." In the External conditions, the potential causes were replaced as "someone sitting next to you" (Controllable) or "a sudden earthquake" (Uncontrollable). In all conditions, the scenario ended with a consequential event: "you found a coffee stain on the book." After reading each scenario, participants were asked to describe how they would explain the situation to the friend. They also rated to what extent they would feel responsible for what happened on a 5-point scale (1 = *not responsible at all*, 5 = *very responsible*).

Table 1

The Experimental Scenarios with the Manipulation of Locus of Causality and Controllability

Senario 1		
BACKGROUND	**Suppose you were reading a book that you borrowed from a friend, while having a coffee.**	
CAUSAL EVENT	(Internal/Controllable)	Your hand hit on the mug
	(Internal/Uncontrollable)	You suddeny became dizzy and your hand hit on the mug
	(External/Controllable)	The hand of someone else sitting next to you hit on the mug
	(External/Uncontrollable)	There suddenly came an earthquake
CONSEQUENTIAL EVENT	**You found a stain of spilt coffee on the book.**	
Senario 2		
BACKGROUND	**Suppose you were in a parking lot.**	
CAUSAL EVENT	(Internal/Controllable)	While parking a car, you looked away for a moment.
	(Internal/Uncontrollable)	While parking a car, you suddeny became dizzy.
	(External/Controllable)	Your friend X was parking a car, but X was not a good driver.
	(External/Uncontrollable)	While you were parking a car, the brake suddeny stopped working.
CONSEQUENTIAL EVENT	**The car contacted with the next car, leaving a scratch on it, which tured out to be owned by your friend Y.**	

As we expected, participants typically provided explanations for a *causal event* (e.g., the coffee mug being tipping over) and a *consequential event* (i.e., stain on the book) in various words, and the use of transitive verbs in these events was the main dependent variable in our analysis. Despite the fact that the scenarios used no transitive verbs, participants spontaneously produced transitive sentences such as "(I) spilled the coffee" more often than not. We coded the presence of transitive verbs as "1" for causal and consequential events, respectively, with the scores

hence ranging from 0 (i.e., typically intransitive verbs for both events) to 1 (transitive verb for either event) to 2 (transitive for both)[2].

The top row of Table 2 shows how this transitivity score varied across the experimental conditions. Look at the first two cells from left combined (i.e., the Internal condition) and compare it to the next two cells combined (External condition). Clearly, the descriptions in the former condition included, on average, a greater number of transitive verbs ($M = 0.72$) than in the second condition ($M = 0.36$). This is the main effect for Locus of Causality (ignoring the other variable, Controllability). An Analysis of Variance (ANOVA) revealed that this difference was statistically significant (see also the right part of Table 2). On the other hand, the manipulation of Controllability did not result in any significant difference ($M = 0.49$ for the two cells under the label, Controllable, combined, vs. $M = 0.58$ for the other two cells, Uncontrollable, combined).

Table 2
Means (and Standard Deviations) of Verb Transitivity and Perceived Responsibility

		Locus				ANOVA Results			
		Internal		External					
		Controllability							
Scenario	Measure	controllable	uncontrollable	controllable	uncontrollable	Effect	$F(1, 54)$	p	η^2
Coffee	Verb Transitivity	0.75	0.69	0.24	0.47	L	30.86	< .001	0.36
		(0.55)	(0.66)	(0.43)	(0.50)	C	1.96	.168	0.03
						L x C	5.31	.025	0.09
	Responsibility	4.93	4.31	2.89	3.25	L	136.38	< .001	0.72
		(0.26)	(0.87)	(1.28)	(1.27)	C	0.55	.463	0.01
						L x C	19.90	< .001	0.27
Parking	Verb Transitivity	0.86	0.78	0.16	0.51	L	55.24	< .001	0.51
		(0.48)	(0.49)	(0.37)	(0.50)	C	5.99	.018	0.10
						L x C	10.96	.002	0.17
	Responsibility	4.96	4.80	2.47	3.84	L	129.53	< .001	0.71
		(0.19)	(0.44)	(1.78)	(1.26)	C	17.33	< .001	0.24
						L x C	26.07	< .001	0.33

- Verb transitivity scores range from 0 to 2.
- Responsibility rating: 1 = not responsible at all; 5 = extremely responsible.
- Effect in ANOVA: L = main effect for Locus, C = main effect for Controllability, L x C = Locus x Controllability interaction

More importantly, however, a significant 2-way interaction revealed that it would be premature to conclude that controllability made no difference. In fact, controllability did influence the use of transitive verbs

2 As already noted, pronouns are often dropped in Japanese. We coded transitive verbs without the sentence subject as "first person transitive." This was reasonable for the three conditions except for External/Controllable because the scenarios described only one human agent, "You." In the External/Controllable condition, responses such as "Someone was sitting next to me, and seems like, (pronoun dropped) knocked over the coffee, and so *stained* your book" could be interpreted either as referring to that "someone" or to the simulated self who was reading the book and having the coffee. We decided to code these cases as "first-person" for the sake of consistency across the different conditions.

depending on the other variable, i.e., Locus. When the cause was internal to the protagonist, which simulated the participants themselves, the transitivity scores did not differ statistically whether the cause was controllable (i.e., carelessness; M = 0.75) or uncontrollable (i.e., sudden dizziness; M = 0.69). On the other hand, the story changed significantly in the External condition. Transitive verbs were more likely to be used when the cause was an uncontrollable earthquake (M = 0.47) than when it was attributable to something controllable by a third person (M = 0.24) (p < .05). Table 2 shows that the pattern found for the rating of perceived responsibility paralleled the one for the transitivity scores.

In the second "parking scenario," the protagonist in the Internal condition either looked away (Controllable) or suddenly became dizzy (Uncontrollable) while parking a car (see the lower portion of Table 1). The External/Controllable condition replaced the driver as "Friend X, a lousy driver, was parking," whereas in the External/Uncontrollable condition, there was a sudden brake failure while "You" were parking. In all scenarios, the car "*had a contact with* [i.e., *sesshoku shita*] another car in the next parking spot," which turned out to be owned by "your friend, Y." Results concerning the transitivity scores and perceived responsibility showed essentially the same pattern as in the previous coffee scenarios. Here too, the 2-way interaction was significant on both measures, and particularly notable was the result that, even when the cause was external and uncontrollable (i.e., brake failure), the verb transitivity and responsibility ratings were both significantly higher than when the cause was relatively controllable (i.e., clumsy third-person driver) (both ps < .001).

Hence, the overall trend across the two scenarios showed the more frequent use of transitive verbs in describing an event of damaging someone else's property, particularly when the assumed cause was internal rather than external to the simulated first person. This was the case even though most of these descriptions in Japanese did not include the sentence subject and thus needed to be inferred. In addition, perceived responsibility showed a similar pattern.

Additionally, the unexpected effect of uncontrollability inviting higher transitivity and responsibility scores should warrant further examination. There are at least two candidate explanations. The first pertains to the already-mentioned linguistic rule in Japanese to prohibit the use of an inanimate object as the subject of a transitive sentence. Specifically, statements such as "An earthquake caused the coffee stain on the book" are grammatically not allowed, and a qualification like "Because there was an

earthquake…" is required. And yet, it is still interesting that a substantial number of participants continued to write, "(pronoun dropped) stained the book," rather than "the book has a stain on it." The same was true for a "sudden brake failure."

The second possibility is that participants might have chosen transitive verbs with a certain level of readiness for potential blame. This interpretation is consistent with the results concerning perceived responsibility. As seen in Table 2, the overall pattern was parallel to the results concerning the use of transitive verbs. Just as the External/Uncontrollable condition showed greater transitivity in word choice, a stronger sense of responsibility was expressed in this condition than in Controllable (p < .01). In other words, when no human agent was identified, the Japanese participants expressed that they would take potential responsibility to a moderate level even if they knew they were not causally at fault. Certainly, the expressed level of perceived responsibility was not as high as in the cases of clearly internal causes, and yet, implying the self-responsibility even to this degree would be unlikely for English speakers. (Remember the previously mentioned anecdote of returning a broken appliance to the landlord.)

Replications and extensions

Arguably, limitations of the preliminary study summarized above should be pointed out, including its solely within-participant design, a limited variation in scenario situations, and the reliance on a Japanese student sample alone. This study, however, stimulated a number of subsequent studies with methodological refinements. Results from these studies generally replicated the original pattern of findings as to Japanese speakers and furthermore provided cross-linguistic evidence. For instance, Yoshinari, Pardeshi, and Chung (2010) added speakers of Marathi (spoken in India) and Korean in their analysis, while using a different hypothetical event ("dropping and breaking a dish at the dinner table"). Here too, verb transitivity among Japanese speakers was constantly high for Internal conditions regardless of controllability, and moderate for the External/Uncontrollable "earthquake" scenario. Korean speakers showed a different pattern, with a relatively low and constant level of transitivity across the conditions. These results suggest that even though Japanese and Korean share a great deal of similarity in their syntactic features, pragmatic roles of transitivity in this specific context (i.e., explaining

property damage) can be somewhat different. (The same pattern of difference was found by Yoshinari, Pardeshi, & Chung, 2012, with a more fine-tuned coding of verbs). In an even more visible contrast, verb transitivity among the Marathi speakers was essentially null in any of the conditions. Marathi hence appears to be an extreme case among the "Indo-European languages" Ikegami (1991) discussed. Even in the context of a close social relationship, speakers of this language never used transitive verbs to explain an unintentional and accidental incident.[3]

The sharp distinction in the use of transitive and intransitive verbs in Marathi was also demonstrated by Pardeshi and Yoshinari (2012). In this study, the authors prepared video clips that were similar to the ones used by Fausey and Boroditsky (2011) and consisted of multiple behavioral segments so that the verb transitivity in open-ended descriptions could range from 0 to 3: namely, (1) contact of the protagonist's hand with the mug, (2) toppling of the cup, and (3) spilling of the liquid. These were presented to Japanese and Marathi speakers based on a between-participant design. Results showed that the differentiated use of transitive verbs for intentional (vs. intransitive for accidental) events was commonly found for *both* language users. However, the differentiation was significantly more pronounced for Marathi than for Japanese (see Figure 1). Hence, even when the judgment was made from a viewpoint of an independent observer, the use of transitive verbs in describing an accidental event indeed remained among Japanese, but the level was lowered as compared to the situations with a greater self-involvement (i.e., the "coffee" and "parking" scenarios used in the studies mentioned earlier). Among Marathi speakers, on the other hand, the use of transitive verbs was relatively enhanced for intentional events, whereas that of intransitive decreased to a zero level.

Focusing more on the implication of social relationships, we designed a different study to examine the morally laden effect of verb transitivity (Karasawa, Suga, & Sato, 2017). Incorporating the "question and answer paradigm" (Semin, 2000), we prepared 16 pairs of scenarios, including interrogations either in transitive or intransitive forms (e.g., "Tell me why (you) broke the window" vs. "… why the window broke").[4] Ratings of

---

3   English speakers use quasi-transitive expressions at times, such as "I *cut* my finger when I was cooking," or "I *broke* my leg," but Marathi normally requires, "I *had* my finger/leg cut/broken."

4   The Japanese interrogative used in this study, *dooshite*, may actually mean "how" as well as "why." This may corroborate Ikegami's (1991) characterization of Japanese in that "In what way?" tends to be highlighted more than "Who done it?" possibly both linguis-

each character in the scenarios revealed that the questioners using transitive rather than intransitive verbs were perceived to be blaming more, a colder (vs. warmer) character, and interestingly, holding a greater social power between the two. Apparently, the pragmatic rule among Japanese speakers dictates that using transitive terms to blame is allowed for someone with a greater but not a lesser power.

The Use of Transitive Verbs for Intentional and Accidental Events
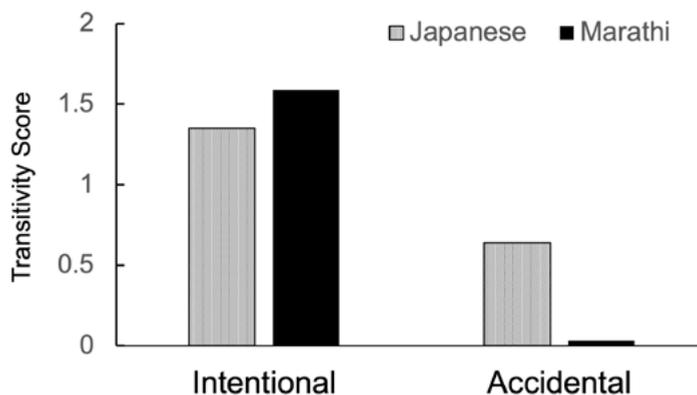
Note. Scores ranged from 0 to 3.
Reproduced from Pardeshi & Yoshinari (2012, p.82) with the authors' permission

Figure 1. Use of transitive verbs

As reviewed above, studies of verb transitivity have significant potential to raise a variety of interesting questions for areas of social and cultural psychology. Psychological research in these areas, including the ones from our own research programs, is still in its preliminary stage, and much more research is needed.[5] Specifically, previous studies have been limited in their range of languages, variations in social contexts that were investigated, and the methodology of typically using hypothetical vignettes.

---

tically and psychologically.
5   An undergraduate research project by Cristiano Zanettti under the supervision by Anne Maass at the University of Padova should be acknowledged as a unique attempt.

Conclusion

Representing agentivity of an event, associated with factors such as causality, volitionality, intentionality, and the directness of influence, by the use of transitive rather than intransitive verbs appears to be a universally observed tendency (Hopper & Tompson, 1980). This linguistic tendency should be well understood by ordinary native speakers of various languages. Above and beyond the universality, however, important differences across languages also exist, such that even the use of the same grammatical form (e.g., transitive verbs) may present different pragmatic meanings. The discrepancy has profound implications, particularly when situated in social relationships. Specifically, verb transitivity may be associated with inferences concerning causality in similar ways among Japanese and English speakers. Nevertheless, the social and psychological consequences of the use of transitive verbs, including the feeling of responsibility and blaming, may vary across linguistic cultures, as our own works reviewed above have demonstrated. The potential difference in the interpretation of the same word choice may give rise to not only misunderstanding but also a more complicated situation of intercultural contact, such as conflicts and contentions between individuals and groups. In contrast to the voluminous body of grammatical analyses conducted by linguists, psychological research investigating the pragmatic meanings of agentivity is still sparse. Collaborative research by scholars of diverse academic disciplines and different linguistic-cultural backgrounds is needed to address possible solutions for those problems.

From a broader theoretical perspective, the present analysis characterized the differential use of transitive and intransitive as a manifestation of a more deeply rooted linguistic difference between the agentivity-oriented DO-languages and the context-oriented BECOME-languages (Ikegami, 1991). Further research may pursue other aspects of the assumed difference. For instance, the strong tendency in English to personify the constituents of a behavioral event and a mental state (e.g., "I don't understand *you*") may reflect in the well-known psychological phenomenon called Spontaneous Trait Inference (STI). That is, when we observe someone engaging in an act (e.g., hitting someone), we normally make instant inferences about a stably enduring trait characteristics of the actor (e.g., "violent" or "aggressive") even when we do not intend to do so (Uleman, Saribay, & Gonzalez, 2008). Linguistically, this "inductive" process of characterizing an act and its actor typically reflects in the frequent use of trait terms like nouns and adjectives rather than more context-dependent

predicates such as verbs in describing behavioral episodes. The opposite direction of inference, namely, a "deduction" is observed less frequently (i.e., the Induction-Deduction Asymmetry, IDA; Maass, Columbo, Columbo, & Sherman, 2001). Empirical evidence has established that both STI and IDA are more prevalent among speakers of DO-languages (e.g., English and Italian) than among those of BECOME-languages (e.g., Japanese and Korean) in their linguistic (e.g., word choice) and cognitive representations (e.g., memory) (Kashima, Kashima, Kim, & Gelfand, 2006; Maass, Karasawa, Politi, & Suga , 2006; Morris & Mok, 2011; Rhee, Uleman, Lee, & Roman, 1995; Shimizu, Lee, & Uleman, 2017).

As retrospect, the Kobe Symposium including Anne Maass as a speaker has opened a path to this new area of research in language and social psychology. Here too, because the "symposium" is an inanimate object, speakers of proper Japanese may only be able to say, "Greatly owing to the panelists, this area of research has begun to flourish," in a literal translation. Yet, by no means should one interpret that the intransitive form diminishes its agentic and causal implication. Instead, as our data on responsibility judgments suggests, they will not forget to add, "(We) respectfully acknowledge their contributions."

References

Bettinsoli, M. L., Maass, A., Kashima, Y., & Suitner, C. (2015).Word-order and causal inference: The temporal attribution bias. Journal of Experimental Social Psychology, 60, 144-149.

DeLancey, S. (1985). Agentivity and syntax. CLS (Chicago Linguistic Society) 21: 1-12.

Fausey, C. M., & Boroditsky, L. (2011). Who dunnit? Cross-linguistic differences in eye-witness memory. Psychonomic Bulletin and Review, 18, 150-157.

Fausey, C. M., Long, B. L., Inamori, A., & Boroditsky, L. (2010). Constructing agency: The role of language. Frontiers in Cultural Psychology, 1, Aiticle 162, 1-11.

Hinds, J. (1986). Situation vs. person focus (with a preface by Yoshihiro Nishimitsu). Kuroshion Publishers.

Heider, F. (1958). The psychology of interpersonal relations. Wiley.

Hopper, P. J. & Thompson, S. (1980). Transitivity in grammar and discourse. Language, 56, 251-99.

Ikegami, Y. (1982). 'Indirect Causation' and 'De-agentivization': The

semantics of involvement in English and Japanese. Proceedings of the Department of Foreign Languages and Literature, College of Arts and Sciences, University of Tokyo, 29-3, 95-112.

Ikegami, Y. (1991). DO-language and BECOME-language: two contrasting types of linguistic representation. In Ikegami, Y. (Ed.). The Empire of Signs: Semiotic Essays on Japanese Culture (pp. 285-326). John Benjamins.

Jacobson, W. (1991). The transitive structure of events in Japanese. Kurosio Publishers.

Karasawa, M., Suga, S., & Sato, A. (2017, July 5-8). The use of transitive verbs elicits the perception of blame and social power [Poster presentation]. The 18th General Meeting of the European Association of Social Psychology. Granada, Spain.

Kashima, Y., & Kashima, E. (1998). Culture and language: The case of cultural dimensions and personal pronoun use. Journal of Cross-Cultural Psychology, 29, 461-486.

Kashima, Y., Kashima, E., Kim, U., & Gelfand, M. (2006). Describing the social world: How is a person, a group, and a relationship described in the East and the West? Journal of Experimental Social Psychology, 42, 388-396.

Malle, B. F., Guglielmo, S., & Monroe, A. E. (2014). A theory of blame. Psychological Inquiry, 25, 147-186

Maass, A., Columbo, A., Columbo, A., & Sherman, S. J. (2001). Inferring traits from behaviors versus behaviors from traits: The induction-deduction asymmetry. Journal of Personality and Social Psychology, 81, 391-404.

Maass, A., Karaawa, M., Politi, F., & Suga, S. (2006). Do verbs and adjectives play different roles in different cultures? A cross-linguistic analysis of person representations. Journal of Personality and Social Psychology, 90, 734-750.

Miyamoto, Y, (2013). Culture and analytic versus holistic cognition: Toward multilevel analyses of cultural influences. Advances in Experimental Social Psychology, 47, 131-188.

Morris, M. W., & Mok, A. (2011). Isolating effects of cultural schemas: Cultural priming shifts Asian-Americans' biases in social perception and memory. Journal of Experimental Social Psychology, 47, 117-126.

Mizutani, O. (1987). Hanashi kotoba to nihonjin: Nihongo no seitai [Spoken languages and the Japanese: The ecology of Japanese]. Soutakusha Publishing.

Nisbett, R. E., Peng, K., Choi, I., & Norenzayan, A. (2001) Culture and

system of thought: Holistic versus analytic cognition. *Psychological Review*, 108, 291-310.

Nishimura, Y. (1997). Agentivity and causation in cognitive linguistics. In K. Yamanaka & T. Ohori (Eds.), *The locus of meaning: Papers in honor of Yoshihiko Ikegami*, (pp. 277-292). Kurosio Publishers.

Pardeshi, P. & Yoshinari, Y. (2012). An investigation into the interaction between intentionality and the use of transitive/intransitive expression: A contrastive study of Japanese and Marathi. *Journal of Japanese Linguistics, 28*, 77-88.

Rhee, E., Uleman, J. S., Lee, H. K., & Roman, R. J. (1995). Spontaneous self-descriptions and ethnic identities in individualistic and collectivistic cultures. *Journal of Personality and Social Psychology*, 69, 142-152.

Semin, G. (2000). Launguage as a cognitive and behavioral structuring resource: Question-answer exchanges. *European Review of Social Psychology*, *11*, 75-104.

Shimizu, Y., Lee, H., & Uleman, J. S. (2017) Culture as automatic processes for making meaning: Spontaneous trait inferences. *Journal of Experimental Social Psychology*, 69, 79-85.

Suitner, C., & Maass, A. (2016). Spatial Agency Bias: Representing people in space. *Advances in Experimental Social Psychology*, 53, 245-301.

Uleman, J. S., Saribay, S. A., & Gonzalez, C.M. (2008). Spontaneous inferences, implicit impressions, and implicit theories. *Annual Review of Psychology*, 59, 329-360.

Weiner, B. (1986). *An attribution theory of motivation and emotion.* Springer-Verlag.

Yoshinari, Y., Pardeshi, P., & Chung, S-Y. (2010). *Hiitotekina dekigoto niokeru tadooshi shiyoo to sekinin ishiki: Nihongo, Kankokugo, Marathigo no jittai tyoosa o tsuujite* [Use of transitive verbs in depicting accidental events and perception of responsibility: Through the survey of actual usages in Japanese, Korean and Marathi]. In H. Kishimoto (Ed.). *Kotoba no taishoo* [Languages in contrast] (pp. 175-189). Kurosio Publishers.

Yoshinari, Y., Pardeshi, P., & Chung, S-Y. (2012). Usage of transitive verbs in the depiction of accidental events in Japanese and Korean. *Japanese/Korean Linguistics, 21*, 231-245.

6. Social Perception of Groups: The Role of Stereotype Content

Federica Durante
University of Milano-Bicocca, Italy

Among the many abilities of our cognitive system, there is the ability to sort the incredible amount of things, animals, people, entities, and events present in our environments into a manageable number of categories. We reach this apparent simplicity through a cognitive process called categorization – when we deal with people: social categorization. This process is at the basis of cognitive biases, schema, prejudice, and, of course, stereotyping.

The simplest way to describe a stereotype is having an idea (no matter if accurate or not) of a social group, which contains traits, characteristics, habits, behaviors, and applying this idea rigidly to all individuals belonging to such a group, regardless of their individual differences. This idea is also very difficult to change since our cognitive system is conservative (Fiske & Taylor, 1984). It can be so difficult that John Bargh (1999) provocatively defined stereotypes as "the cognitive monster," meaning something we will never get rid of. Certainly, this is an extreme provocation, and decades of literature found a few strategies that enable us to mitigate the use of stereotypes in social perception; nevertheless, depending on the context, we apply them automatically.

Social cognition research has focused for a long time on stereotyping processes to uncover the systematic principles underlying stereotypes that could be generalized across contexts. At the beginning of the current century, however, we witnessed a renewed interest in the content of stereotypes. In particular, the Stereotype Content Model (SCM; Fiske et al.,

2002) became extremely popular, fostering research on this issue for the last two decades.

In this chapter, I will first illustrate the SCM and then discuss some moderators of the content of stereotypes. I will conclude with the most recent debate in the literature concerning the basic dimensions characterizing social perception.

The Stereotype Content Model

"Not all stereotypes are alike" (Fiske et al., 2002, p. 878). In other words, despite the cognitive processes underlying the formation of stereotypes can be generalized to all individuals, the stereotypical characteristics associated to a social group may vary depending on the group under scrutiny, the societal culture, and the historical time we are living. Therefore, if on the one hand, we may describe Italians as passionate but disorganized, on the other hand, Germans are stereotypically described as well-organized but cold. These, and many others, characteristics represent the stereotype content. But are there underlying, basic dimensions (i.e., systematic kinds of impressions; Fiske, 2018) that could be found in all stereotype contents across contexts? According to the SCM, the answer is yes, there are: warmth and competence.

Warmth refers to traits such as trustworthiness, friendliness, sociability, and sincerity, whereas competence reflects ability, intelligence, skillfulness, and capacity. Forming a quick impression of people is fundamental for human survival. We need to know if we can trust (or not) the individuals we interact with, and we must understand their ability to enact, in case, bad intentions. This evolutionistic approach, as well as the functionalistic one of "thinking is for doing" (Fiske, 1992), drive the SCM to maintain that warmth and competence are 'fundamental' in group perception, representing the stereotype content's core.

Bearing in mind that the SCM focuses on societal stereotypes, namely known, shared, cultural views of groups,[1] Fiske and colleagues (2002) argued that (known) socio-structural relations between groups, within

---

1 The SCM method includes first a preliminary study in which participants are asked to list the most salient social categories in their society. Then, the most mentioned categories are selected, and new samples are asked to evaluate such groups on SCM traits, on Likert's scales, according to the *society's point of view*. This approach allows the assessment of stereotype contents of relevant societal categories while avoiding desirability issues since participants are not asked about their own beliefs.

a society, allow predicting the groups' warmth and competence ascriptions. Having (vs. lacking) status and power lead to attributions of high (vs. low) competence; and being cooperative (vs. competitive)[2] predicts attributions of high (vs. low) warmth. Because status and interdependence (i.e., cooperation/competition) can be high vs. low, as well as the following ascriptions of competence and warmth, the SCM conceptualized four types of stereotype content. The SCM taxonomy named such types in terms of the emotions that the content elicits in perceivers (i.e., prejudice): admiration (high competence and warmth); contemptuous (low competence and warmth); paternalistic (low competence but high warmth); envious (high competence but low warmth). Paternalistic and envious stereotypes, because of their mixed combination of warmth and competence (high on one dimension, low on the other), are defined ambivalent stereotypes.

The SCM initial claim was that most societal stereotypes are ambivalent. Although later studies showed that it was not necessarily true for all contexts (e.g., US vs. European countries), the concept of ambivalent evaluations of groups on warmth and competence allowed the discipline to move a step further from the idea of prejudice as just an overgeneralized antipathy (even hate) towards groups. The concept of positive stereotypes was already in the field (see Czopp et al., 2015 for a review), as well other theoretical attempts to account for modern forms of prejudice and stereotyping (e.g., Gaertner & Dovidio, 1986; Kats & Hass, 1988; McConahay, 1986; Pettigrew & Meertens,1995). The changes in (at least) Western societies in terms of social norms that has begun in the last century, more and more sanction the overt expression of prejudice. This has challenged scholars to find new methodological ways to measure (e.g., implicit measures) and conceive contemporary prejudice. The SCM gave an additional boost to this evolution in the field of stereotyping and prejudice by introducing a relatively simple model that, starting from their predictors, encompasses the content of groups' stereotypes, the related emotions/prejudice, and (later) the linked behavioral tendencies (see the BIAS map; Cuddy et al., 2007).

Most importantly, the SCM argued how ambivalence may be a form of societal control. As noted by Kay and colleagues (2007), "the content of

---

2 Group's competition/cooperation can be measured in terms of economic interests (opposite vs. convergent) and symbolic values (conflicting vs. shared). Studies showed that the best assessment should include both (Kervyn et al., 2015), consistent with the realistic and symbolic theoretical framework (Stephan et al., 2009).

stereotypes can serve to maintain ideological support for the prevailing social system" (p. 312). Because stereotypes function to justify the status quo (Fiske, 1993; Jost & Banaji, 1994; Jost, Burgess, & Mosso, 2001; Tajfel, 1981), having something positive to say about groups, either on warmth or competence, not only may serve to justify the position of groups into the societal hierarchy (e.g., poor people lack competence), but also camouflage the negative side (lacking competence but warm). Blatant antipathy toward a group is more easily recognizable by its targets and, therefore, it can be challenged, resisted, fought back. Ambivalence, instead, flatters social groups on one dimension, bypassing groups' resistance (see Durante et al., 2013). Additionally, the research conducted so far (both within and outside the SCM framework) suggests that the positive characteristics contribute, like the negative ones, to perpetuate the differences between groups in maintaining the status quo and the privilege of some groups over others (see Czopp, et al., 2015). For instance, in communication is common to neglect negative information about social targets while highlighting the positive ones (i.e., stereotyping by omission). However, a phenomenon known as the innuendo effect takes place when warmth and competence are concerned: when only positive characteristics of either warmth or competence are mentioned, the listeners tend to infer the negative aspects of the complementary dimension that is omitted (Bergsieker et al., 2012). Hence, the stereotype is perpetuated even when not overtly displayed. Consistently, the compensation-hypothesis studies (compensation effect; Judd, et al., 2005) revealed trade-offs of competence and warmth when people rate individuals or groups in a comparative context. Namely, a compensatory process occurs when a positive perception on one dimension is offset by a negative perception on the other (see also, Kervyn et al., 2010, 2016).

Cultural, Social, and Economical Moderators of the Stereotype Content

As said, the SCM fostered two decades of research. Despite the SCM being generally supported, cultural variations were observed (see Fiske & Durante, 2016 for an overview). In the US society (e.g., Bergsieker et al., 2012; Cuddy et al., 2007; Fiske et al., 2002), for instance, it is usual to find ingroups, allies, reference groups (e.g., middle class) in the admiration cell of the model (i.e., being perceived as high on both dimensions); immigrant and homeless people in the contemptuous cell (i.e., low on

both dimensions); elderly and children in the paternalistic one (i.e., high warmth but low competence); rich people and professionals in the envious cell (i.e., high competence but low warmth). However, in a cross-cultural study, Cuddy and colleagues (2009), by comparing individualistic (European) versus collectivistic (East Asian) countries, found that, in the latter countries, ingroups and reference groups did not appear in the admiration cell, receiving instead middle evaluations on both dimensions. The authors argued that collectivistic norms (e.g., modesty, humility, and harmony) moderated the positive attributions of such groups.

Durante et al. (2013) considered data from 25 different nations (at least one sample from each continent), and found another moderator of the stereotype content, namely economic inequality (measured by the Gini Index). In the US and other countries with moderate to high income inequality (Latin America, South Africa) many salient, social groups were stereotyped ambivalently. In contrast, in more egalitarian countries, such as Australia, and much of Europe, groups tended to be perceived as "all-good insiders" (high on both dimensions) or "all-bad outsiders" (low on both dimensions) (Fiske, 2018). This suggests that in countries with more economic equality (likely, less competition), there is a larger inclusive ingroup–"us"–versus few outcasts–"them." Income inequality, instead, creates more complicated societal situations (e.g., frictions between groups, perceived unfairness) that require explanations. To this aim, as discussed above, ambivalent portrayals of groups may help societies to justify large economic disparities (e.g., deserving and culpable poor, meriting and undeserving rich; see Durante et al., 2017).

If ascriptions of warmth and competence are driven by the survival need to understand and distinguish friends from foes, what happens to the content of stereotypes when conflicts (e.g., wars) characterize a society? Durante, Fiske et al. (2017) collected SCM data from very peaceful (e.g., Denmark) and high-conflict (e.g., Iran) countries and adding such data to the Durante et al.'s (2013) database. They considered the Global Peace Index (GPI) and, again, the Gini Index as moderators of the level of ambivalent stereotype content. Results showed that the pattern described above for egalitarian countries also characterized peaceful countries as well as high-conflict countries, indicating a simple "us vs. them" dynamic (less ambivalent stereotypes). Most interestingly, the countries intermediate on the GPI (e.g., the US) displayed the clearest pattern of stereotype ambivalence. Hence, the study showed a curvilinear relationship between stereotype ambivalence and conflict. The authors reasoned that very

peaceful countries likely need less ambivalence because most groups fit in the shared national identity, with few outcasts (as for inequality); but also high-conflict societies need a clearcut distinction between friends and foes to simplify the images of groups; on the contrary, nations with intermediate conflict are likely characterized by ambiguous intergroup relations, neither entirely peaceful nor entirely conflictual, and may need ambivalence for system stability.

Warmth and Competence: Is that all there is to the Stereotype Content?

While many SCM studies were carried out, an important scientific debate on the dimensions underlying social perception emerged, generating new research and theoretical frameworks.

Warmth and competence, also called the "Big Two" of social perception (Abele & Wojciszke, 2014), have long been acknowledged by social psychology, although in person perception. We can trace them back to Asch's (1946) impression formation experiments and, tested more formally, to Rosenberg et al. (1968) and Zanna and Hamilton (1972). However, similar 'fundamental' dimensions appeared in the literature. The most known ones are agency and communion (Bakan, 1966). These dimensions were investigated in particular   in the context of self-concept (Abele et al., 2016) and interpersonal attitudes (Wojciszke et al., 2009). Agency includes attributes considered relevant to reaching goals. Communion, instead, refers to attributes relevant to   social relationships. Agency and communion both entail two subdimensions (or facets): the former can be distinguished in assertiveness and ability (originally labeled as competence; Abele et al., 2016); the latter includes friendliness (originally labeled as warmth; Abele et al., 2016) and morality.

Building on the agency and communion literature, other streams of studies distinguished competence, sociability, and morality, and claims that morality (i.e., "a facet of warmth that comprises characteristics relevant to perceived correctness of social targets"; Brambilla et al. 2011, p. 136), over competence and sociability, is the most important dimension in forming impressions of individuals and groups (e.g., Brambilla et al., 2011, 2021; Brambilla & Leach, 2014), or the most important dimension in evaluating ingroups (e.g., Ellemers, 2017; Leach et. al, 2007). Yzerbyt and colleagues (e.g., Kervyn at al., 2010; Yzerbyt, 2018) took on the warmth and competence dimensions as conceptualized by the SCM but

focused on the way they interact when forming impressions in comparative contexts, individuating the compensation effect (discussed above). On the contrary, the Agency-Beliefs-Communion (ABC) Model (Koch et al., 2016), took a different approach: using a data-driven, rather than a theory-driven, method, claims that when people form spontaneous impressions of groups, they organize them by evaluating groups in terms of how similar they seem. In doing that, they use three dimensions: agency (which includes status) beliefs (conservative vs. progressive), and communion.

Therefore, the literature seems to converge on considering warmth/communion and competence/agency as the underlying dimensions of social perception, but with the caveat of acknowledging their sub-dimensions, namely sociability and morality for the former, and ability and assertiveness for the latter. To reconcile with the beliefs dimension brought out by the ABC model, the researchers have started a so-called adversarial collaboration (see Ellemers et al., 2020). First, they distinguished their models from a theoretical point of view,[3] highlighting convergent and divergent evidence for each model and its predictions. Then, they suggested an integrated framework, arguing about two fundamental dimensions, namely, vertical (agency, competence) and horizontal (communion, warmth)–and their respective sub-dimensions–and including beliefs as a context-dependent third dimension, which may emerge in social perception of groups depending on the perceivers' goals, domains, situations. Finally, they argued that the methodological approach (e.g., data-driven vs. theory-driven) may play a role in terms of the prevalence of one facets/dimensions over the others (Abele et al., 2021).

A recent work, with a methodological pick, investigated spontaneous stereotype contents (Spontaneous Stereotype Content Model; SSCM; Nicolas et al., 2022a). The authors aimed to investigate the most common dimensions spontaneously mentioned by people when judging salient social groups. For this purpose, they used natural language processing text analyses, tested the robustness of the dimensions that emerged, and then contrasted these findings with those from other methodological approaches (i.e., Likert scales and open-ended questions). Results showed

---

[3] In this adversarial collaboration, studies by Abele, Wojciszke and colleagues were organized into the Dual Perspective Model (DPM); studies by Ellemers and colleagues into the Behavioral Regulation Model (BRM); studies by Yzerbyt and colleagues into the Dimensional Compensation Model (DCM) (Abele et al., 2021). Recently, also studies by Brambilla and colleagues were organized into the Moral Primacy Model (MPM; Brambilla et al., 2021).

the importance of considering the facets of warmth and competence (as indicated above), as they can vary independently depending on the societal group. Also, beliefs appeared as a relevant dimension but in a less prevalent way. Finally, they uncovered other dimensions not considered previously such as emotions, health, appearance, and others. The prevalence of one dimension/facets over the other however was further investigated in light of the perceivers' current goal (Nicolas et al., 2022b). Results showed that participants prioritized learning about targets' sociability and especially morality (i.e., warmth) across conditions. However, having a relational goal (gathering information about an unknown group entering one's neighborhood) versus a structural one (unknown immigrant group) can moderate priorities, namely an increased interest in the sociability facet versus beliefs and competence, respectively.

To conclude, stereotype content matters but its underlying dimensions, despite some agreements, are still debated. This is not unusual in scientific disciplines. On the contrary, discrepancies in findings fuel the progress of science. There is more to come.

References

Abele, A. E. & Wojciszke, B. (2014). Communal and agentic content in social cognition: A Dual Perspective Model. Advances in Experimental Social Psychology, 50, 195-255. https://doi.org/10.1016/B978-0-12-800284-1.00004-7

Abele, A. E., Ellemers, N., Fiske, S. T., Koch, A., & Yzerbyt, V. (2021). Navigating the social world: Toward an integrated framework for evaluating self, individuals, and groups. Psychological Review, 128, 290–314. https://doi.org/10.1037/rev0000262

Abele, A. E., Hauke, N., Peters, K., Louvet, E., Szymkow, A., & Duan, Y. (2016). Facets of the fundamental content dimensions: Agency with competence and assertiveness— Communion with warmth and morality. Frontiers in Psychology, 7, 1810. https://doi.org/10.3389/fpsyg.2016.01810

Asch, S. E. (1946). Forming impressions of personality. The Journal of Abnormal and Social Psychology, 41(3), 258–290. https://doi.org/10.1037/h0055756

Bakan, D. (1966). The Duality of Human Existence. An Essay on Psychology and Religion. Chicago, IL: Rand Mcnally.

Bargh, J. A. (1999). The cognitive monster: The case against the controllability

of automatic stereotype effects. In S. Chaiken & Y. Trope (Eds.), Dual-process theories in social psychology (pp. 361–382). The Guilford Press.

Bergsieker, H. B., Leslie, L. M., Constantine, V. S., & Fiske, S. T. (2012). Stereotyping by omission: Eliminate the negative, accentuate the positive. Journal of Personality and Social Psychology, 102(6), 1214–1238. https://doi.org/10.1037/a0027717

Brambilla, M., & Leach, C. W. (2014). On the importance of being moral: The distinctive role of morality in social judgment. Social Cognition, 32, 397–408. https://doi.org/10.1521/soco.2014.32.4.397

Brambilla, M., Rusconi, P., Sacchi, S., & Cherubini, P. (2011). Looking for honesty: The primary role of morality (vs. sociability and competence) in information gathering. European Journal of Social Psychology, 41, 135–143. https://doi.org/10.1002/ejsp.744

Brambilla, M., Sacchi, S., Rusconi, P., & Goodwin, G. P. (2021). The primacy of morality in impression development: Theory, research, and future directions. In B. Gawronski (Ed.), Advances in experimental social psychology (Vol. 64, pp. 187–262). Academic Press. https://doi.org/10.1016/bs.aesp.2021.03.001

Cuddy, A. J. C., Fiske, S. T., & Glick, P. (2007). The BIAS map: Behaviors from intergroup affect and stereotypes. Journal of Personality and Social Psychology, 92, 631–648. https://doi.org/10.1037/0022-3514.92.4.631

Cuddy, A. J. C., Fiske, S. T., Kwan, V. S. Y., Glick, P., Demoulin, S., Leyens, J. P., Bond, M. H., Croizet, J. C., Ellemers, N., Sleebos, E., Htun, T. T., Kim, H. J., Maio, G., Perry, J., Petkova, K., Todorov, V., Rodríguez-Bailón, R., Morales, E., Moya, M., … Ziegler, R. (2009). Stereotype content model across cultures: Towards universal similarities and some differences. British Journal of Social Psychology, 48(1), 1-33. https://doi.org/10.1348/014466608X314935

Czopp, A. M., Kay, A. C., & Cheryan, S. (2015). Positive stereotypes are pervasive and powerful. Perspectives on Psychological Science, 10(4), 451–463. https://doi.org/10.1177/1745691615588091

Durante, F., Fiske, S. T., Gelfand, M., Crippa, F., Suttora, C., Stillwell, A., Asbrock, F., Aycan, Z., Bye, H. H., Carlsson, R., Björklund, F., Daghir, M., Geller, A., Larsen, C. A., Abdel Latif, A., Mähönen, T. A., Jasinskaja-Lahti, I., & Teymoori, A. (2017). Ambivalent stereotypes link to peace, conflict, and inequality across 38 Nations. Proceedings of the National Academy of Sciences of the United States of America (PNAS), 114(4), 669-674. https://doi.org/10.1073/pnas.1611874114

Durante, F., Fiske, S. T., Kervyn, N., Cuddy, A. J. C., Akande, A., Adetoun, B. E., Adewuyi, M. F., Tserere, M. M., Al Ramiah, A., Mastor, K. A.,

Barlow, F. K., Bonn, G., Tafarodi, R. W., Bosak, J., Cairns, E., Doherty, C., Capozza, D., Chandran, A., Chryssochoou, X., … Storari, C. C. (2013). Nations' income inequality predicts ambivalence in stereotype content: how societies mind the gap. British Journal of Social Psychology, 52(4), 726-746. https://doi.org/10.1111/bjso.12005

Durante, F., Tablante, C. B., & Fiske, S. T. (2017). Poor but warm, rich but cold (and competent): Social classes in the stereotype content model. Journal of Social Issues, 73(1), 138–157. https://doi.org/10.1111/josi.12208

Ellemers, N. (2017). Morality and the regulation of social behavior. Milton Park, UK: Routledge/Taylor & Francis. http://dx.doi.org/10.4324/9781315661322

Ellemers, N., Fiske, S. T., Abele, A. E., Koch, A., & Yzerbyt, V. (2020). Adversarial alignment enables competing models to engage in cooperative theory building toward cumulative science. Proceedings of the National Academy of Sciences, 117(14), 7561-7567. https://doi.org/10.1073/pnas.1906720117

Fiske, S. T. (1992). Thinking is for doing: Portraits of social cognition from daguerreotype to laserphoto. Journal of Personality and Social Psychology, 63(6), 877–889. https://doi.org/10.1037/0022-3514.63.6.877

Fiske, S. T. (1993). Controlling other people. The impact of power on stereotyping. American Psychologist, 48, 621–628. https://doi.org/10.1037/0003-066X.48.6.621

Fiske, S. T. (2018). Stereotype content: Warmth and competence endure. Current Directions in Psychological Science, 27(2), 67–73. https://doi.org/10.1177/0963721417738825

Fiske, S. T., Cuddy, A. J., Glick, P., & Xu, J. (2002). A model of (often mixed) stereotype content: Competence and warmth respectively follow from perceived Status and competition. Journal of Personality and Social Psychology, 82(6), 878–902. https://doi.org/10.1037/0022-3514 .82.6.878

Fiske, S. T., & Durante, F. (2016). Stereotype content across cultures: Variations on a few themes. In M. J. Gelfand, C.-Y. Chiu, & Y.-Y. Hong (Eds.), Handbook of advances in culture and psychology (Vol. 6, pp. 209–258). New York, NY: Oxford University Press.

Fiske, S. T., & Taylor, S. E. (1984). Social cognition. Reading, MA: Addison-Wesley.

Gaertner, S. L., & Dovidio, J. F. (1986). The aversive form of racism. In J. F. Dovidio & S. L. Gaertner (Eds.), Prejudice, discrimination, and racism (pp. 61–89). Orlando, FL: Academic Press.

Jost, J. T., & Banaji, M. R. (1994). The role of stereotyping in system justification and the production of false consciousness. British Journal

of Social Psychology, 33, 1–27. https://doi.org/10.1111/j.2044-8309.1994.tb01008.x

Jost, J., Burgess, D., & Mosso, C. (2001). Conflicts of legitimation among self, group, and system: The integrative potential of system justification theory. In T. J. Jost, & B. Major (Eds.), The psychology of legitimacy: Emerging perspectives on ideology, justice, and intergroup relations (pp. 363-388). Cambridge: Cambridge University Press.

Judd, C. M., James-Hawkins, L. J., Yzerbyt, V., & Kashima, Y. (2005). Fundamental dimensions of social judgment: Understanding the relations between judgments of competence and warmth. Journal of Personality and Social Psychology, 89, 899–913. https://doi.org/10.1037/0022-3514.89.6.899

Kay, A. C., Jost, J. T., Mandisodza, A. N., Sherman, S. J., Petrocelli, J. V., & Johnson, A. L. (2007). Panglossian ideology in the service of system justification: How complementary stereotypes help us to rationalize inequality. In M. P. Zanna (Ed.), Advances in experimental social psychology, Vol. 39, pp. 305–358). Elsevier Academic Press. https://doi.org/10.1016/S0065-2601(06)39006-5

Katz, I., & Hass, R. G. (1988). Racial ambivalence and American value conflict: Correlational and priming studies of dual cognitive structures. Journal of Personality and Social Psychology, 55, 893-905. https://doi.org/10.1037/0022-3514.55.6.893

Kervyn, N., Bergsieker, H. B., Grignard, F., & Yzerbyt, V. Y. (2016). An advantage of appearing mean or lazy: Amplified impressions of competence or warmth after mixed descriptions. Journal of Experimental Social Psychology, 62, 17–23. https://doi.org/10.1016/j.jesp.2015.09.004

Kervyn, N., Fiske, S. T., & Yzerbyt, Y. (2015). Foretelling the primary dimension of social cognition: Symbolic and realistic threats together predict warmth in the stereotype content model. Social Psychology, 46, 36–45. https://doi.org/10.1027/1864-9335/a000219

Kervyn, N., Yzerbyt, V. Y., & Judd, C. M. (2010). Compensation between warmth and competence: Antecedents and consequences of a negative relation between the two fundamental dimensions of social perception. European Review of Social Psychology, 21, 155–187. https://doi.org/10.1080/13546805.2010.517997

Koch, A., Imhoff, R., Dotsch, R., Unkelbach, C., & Alves, H. (2016). The ABC of stereotypes about groups: Agency/socioeconomic success, conservative-progressive beliefs, and communion. Journal of Personality and Social Psychology, 110, 675–709. http://dx.doi.

org/10.1037/ pspa0000046

Leach, C. W., Ellemers, N., & Barreto, M. (2007). Group virtue: The importance of morality (vs. competence and sociability) in the positive evaluation of in-groups. Journal of Personality and Social Psychology, 93, 234–249. http://dx.doi.org/10.1037/0022-3514.93.2.234

McConahay, J.B. (1986). Modern racism, ambivalence, and the modern racism scale. In J.F Dovidio e S.L. Gaertner (A cura di), Prejudice, discrimination and racism (pp. 91-125). New York: Academy Press.

Nicolas, G., Bai, X., & Fiske, S. T. (2022a). A spontaneous stereotype content model: Taxonomy, properties, and prediction. Journal of Personality and Social Psychology. Advance online publication. https://doi.org/10.1037/pspa0000312

Nicolas, G., Fiske, S. T., Koch, A., Imhoff, R., Unkelbach, C., Terache, J., Carrier, A., & Yzerbyt, V. (2022b). Relational versus structural goals prioritize different social information. Journal of Personality and Social Psychology, 122(4), 659–682. https://doi.org/10.1037/pspi0000366

Pettigrew, T.F. e Meertens, R.W. (1995). Subtle and blatant prejudice in western Europe. European Journal of Social Psychology, 25, 57-75. https://doi.org/10.1002/ejsp.2420250106

Rosenberg, S., Nelson, C., & Vivekananthan, P. S. (1968). A multidimensional approach to the structure of personality impressions. Journal of Personality and Social Psychology, 9, 283-294. https://doi.org/10.1037/h0026086

Stephan, W. G., Ybarra, O., & Morrison, K. R. (2009). Intergroup threat theory. In T. D. Nelson (Ed.), Handbook of prejudice, stereotyping, and discrimination (pp. 43-60). New York, NY, USA: Psychology Press.

Tajfel, H. (1981). Human groups and social categories. Cambridge: Cambridge University Press.

Zanna, M. P., & Hamilton, D. L. (1972). Attribute dimensions and patterns of trait inferences. Psychonomic Science, 27(6), 353–354. https://doi.org/10.3758/BF03328989

Wojciszke, B., Abele, A. E., & Baryla, W. (2009). Two dimensions of interpersonal attitudes: Liking depends on communion, respect depends on agency. European Journal of Social Psychology, 39, 973–990. http://dx.doi.org/10.1002/ejsp.595

Yzerbyt, V. Y. (2018). The dimensional compensation model: Reality and strategic constraints on Warmth and Competence in intergroup perceptions. In A. E. Abele & B. Wojciszke (Eds.), The agency-communion framework (pp. 126–141). London, UK: Routledge. http://dx.doi.org/10 .4324/9780203703663-11

7. Social Grammar Model - how verbs change the wor(l)d

Magdalena Formanowicz[1] and Caterina Suitner[2]
[1]SWPS University, Warsaw, Poland
[2]University of Padova, Italy

> "One thing about which fish know exactly nothing is water, since they have no anti-environment which would enable them to perceive the element they live in."
>
> Marshall McLuhan, War & Peace in the Global Village (1968).

Just as fish know nothing about water, often social scientists are ignorant of language, which is involved in multiple, if not most, of the phenomena we study. Neither beliefs nor attitudes are secretly instilled in people's minds. Similarly, social influence, persuasion, or collective action do not emerge in a vacuum but in the context of words, speeches, talks, communication, and listening (or not listening) to one another. Navigating the social world requires action coordination, for which a shared action representation is needed, and, accordingly, the ability to form shared representations of tasks has been considered a "cornerstone of social cognition" (Sebanz et al., 2006; p. 73; Semin & Cacioppo, 2008; see also Semin, 2000). For the formation of shared representations, language is an indispensable toolkit (Fiedler, 2008), yet social scientists often neglect even what language can do to their measurement. Some studies alarm that the change in a linguistic form used in a questionnaire can alter the results obtained on psychological scales (Vainapel et al., 2015), yet these voices cannot be considered mainstream. In this chapter, we would like to introduce a few exceptions to this rule, proposing that language can represent and affect a more extensive scope of phenomena than previously thought. In the attempt to get a bird's eye view of how language is related

to social phenomena, these few perspectives started with the building blocks of language, linguistic categories. We will specifically focus on the Social Grammar Model (SGM; Formanowicz et al., 2017), which proposes that verbs imply dynamic properties that other grammatical categories (nouns and adjectives) lack, making them the preferred syntactic device to convey activity—and by extension— also social agency, an essential dimension of human perception related to goal achievement (for an overview, see Abele & Wojciszke, 2014). Before we get into describing SGM and its consequences in more detail, we will first review the important intellectual predecessors of SGM. We will conclude the chapter with a brief review of other linguistic tools related to expressing agency and speculate why there are so many cues about agency.

Linguistic Category Model and Linguistic Intergroup Bias

A foundational theory for linguistic social cognition is the Linguistic Category Model (LCM; Semin & Fiedler, 1991; 1988; for the subsequent developments of the model, see: Carnaghi et al., 2008), which established how linguistic categories affect information processing and contribute to representing social information (Fiedler, 2008; Semin, 2008). Specifically, LCM posits that linguistic categories represent information according to its level of abstraction. At the most concrete level, events are construed using descriptive action verbs such as "she kicks the tree". This is considered concrete because the action of kicking unquestionably is tied to the movement of a leg. Therefore, representations built on that sentence will converge across language users as everyone will imagine this action similarly. Interpretative action verbs are considered higher on the abstraction continuum because representations based on them vary greatly. In response to "she hurts the tree", one could imagine not only kicking but also an action of carving a heart with a caption that love will last forever or even signing a petition to cut the tree down. State verbs evoke progressively higher levels of abstraction "she hates the tree", adjectives "she is aggressive", and nouns "she is a vandal". Research conducted in the tradition of LCM established that as the level of abstraction increases, people move from representing actions as an incidental, imaginable, context-dependent, and malleable behavior to a generalized, unspecific, decontextualized, and stable disposition (Semin & Fiedler, 1988; 1991).

Anne Maass with colleagues discovered the potential of the model for the domain of social dynamics (1989, 1995, 1996). The team applied the

LCM to intergroup relations to test whether members of one's own group (ingroup) and of other groups (outgroups) can be portrayed differently and under which conditions that would happen. Given that the ingroup is usually presented in positive (and the outgroup in negative) terms, Linguistic Intergroup Bias (LIB; Maass et al., 1989) combined the analysis of linguistic abstractness with the valanced context. To maximize the generalizability of positive associations, the ingroup is represented more abstractly when describing positive qualities and more concretely when describing negative qualities, as in the sentence "we are honest(abstract positive), even if we did not pay taxes(concrete negative)". In reverse, to stabilize negative associations with an outgroup, undesirable qualities are described by using abstract language, and concrete references are used for positive aspects "they are dishonest(abstract negative), even if they paid taxes (concrete positive)" (Maass et al., 1989; 1995; 1996).

Two mechanisms are proposed to explain the LIB; the first is more motivational, and the second is more cognitive. According to the first motivational account, LIB has the function of enhancing the ingroup (Maass et al., 1996). As we are motivated to build a positive self view, a strategic use of language that advantages the group we belong to (Tajfel & Turner, 1979) fulfills this motivation by describing positive qualities in a generalized way and downplaying negative qualities through the use of concrete language. Furthermore, this allows also for better distinctiveness, namely differentiating ingroup from outgroup: precisely when one's identity is threatened (Brewer, 1979; Brown & Zagefka, 2005), the need for an (also linguistically marked) division between them and us is heightened.

The second explanatory mechanism taps into the cognitive aspects of LIB (Maass et al., 1995). The ingroup is usually represented in positive abstract terms because the positive qualities represent the most frequent references to the ingroup and therefore become generalized in language. As the negative features of an ingroup are usually seen as rare and occasional, they remain on that incidental level and are referenced more concretely - corresponding to their frequency. In general, people tend to describe expected and well-known phenomena in abstract terms "he is always busy" and unexpected information is represented more in terms of specific behaviors "he cooked a surprise dinner" (this phenomenon is also known as the Linguistic Expectancy Bias - LEB; Maass et al., 1995; Wigboldus et al., 2000). By analogy, the same pattern applies to outgroups, for which we expect and thus express more abstractly the negative associations. Positive events are more unexpected and therefore are expressed

more concretely. Importantly, research has found numerous evidence for LIB occurring both in observed (e.g., Dragojevic et al., 2017; Gorham, 2006) and experimental settings (e.g., Assilaméhou & Testé, 2013; Ruscher & Tipler, 2018), establishing the mechanisms operating behind it (for a review, see Maass et al., 2014). Overall, LCM and subsequent models (LIB and LEB) form an excellent example of language strategies that can contribute to forming the shared representation of reality. Abstraction is not only encoded in language through different linguistic categories but can also be successfully decoded from language, with significant consequences on the opinions and judgments of the person receiving the linguistic input.

Capitalizing on the seminal work with the LCM and LIB, we can further refine our understanding of the role of language in cognition by wondering whether the observed effects are driven by the semantic or the grammatical component, which are not distinguished in the LCM. On top of being verbs, adjectives, and nouns, words used in the LCM studies differ in their meaning. Importantly, for the semantic representation, concreteness is considered a fundamental organizing principle (e.g., Feldman et al., 2006; Kousta et al., 2011) based on imaginable (Paivio, 1991) or context-specific experiential (Schwanenflugel & Stowe, 1989) qualities. It is therefore unknown what drives the effects attributed to linguistic categories, whether it is grammar or different meanings related to imaginability or non-grammar related abstractness. This applies to LCM in the following ways. First, to distinguish between verbs, we need to understand and inspect the meaning of a word, because grammatical category is equivalent for all LCM verb types. To distinguish an action verb (kick) from a state verb (hate), meaning is essential, whereas grammar is useless (they are both verbs). Second, LCM does not fine tune abstraction in every class: adjectives can be both concrete "green" and abstract "lovely". Finally, recent norms of concreteness developed by Brysbaert and colleagues (2014), indicate that words of similar meaning, yet belonging to different linguistic or grammatical categories are not necessarily ordered in their level of abstractness as predicted by the LCM. For example, the word "alienate" (interpretative action word - mean concreteness of 1.83) is less concrete than the word "alien" (adjective or noun - mean concreteness of 3.52). Accordingly, we propose that grammatical categories (i.e., verbs vs. adjectives vs. nouns) per se do not carry information about concreteness.

This meaning-based discrepancy was addressed in a set of studies in which experimental stimuli were equalized in terms of their semantics

(Gelman & Heyman, 1999; Walton & Banaji, 2004). Still, "a carrot eater" was seen as more stable in their eating preference than a person "eating carrots." These results were furthermore taken to the realm of affecting behaviors so that "a voter" (label referring to identity) was found to affect more a voting behavior than a reference to voting as an action (Bryan et al., 2011). However, these results were not replicated (Gerber et al., 2016; 2017; Witkowska et al., 2024), which suggests that on top of identity-driven processes, there can be other mechanisms that hinder the effect. One of such processes can be related to the fluency of information processing (Oppenheimer & Frank, 2008) because the use of agent nouns (i.e., nouns that are derived from verbs) is much less frequent than the use of verbs, as indicated by corpora analysis (Formanowicz et al., 2017). Encountering rare stimuli induces salience which in turn may lead to a higher reaction to agent nouns rather than agent verbs. Furthermore, some of the identity processes theorized in the original studies can be nullified by the opposite tendency: verbs evoking higher action tendency (Formanowicz et al., 2017; Foster-Hanson et al., 2020; Rhodes et al., 2019), which brings us to the Social Grammar Model.

Social Grammar Model as a way to express agency

Before we explain why verbs are likely candidates for expressing agency, we would like to highlight the importance of agency to social cognition. Agency refers to goal orientation and the ability to plan and execute goal achievement (Bakan, 1966; Bandura, 2001). Accordingly, agency is important to oneself, because for everyone, achieving their own goal is crucial (Abele & Wojciszke, 2007). Another definition highlights another aspect of agency as the "socioculturally mediated capacity to act" (Ahearn, 2001; p. 112) because goal attainment also depends on contextual factors. This second definition positions agency in the center of social dynamics, in line with the idea that we pay attention to the agency of others because they can facilitate or hinder one's goal pursuit (New et al., 2007). Specifically, research has found that we are attuned to signals of biological motion (Simion et al., 2008), causality, and intentionality (Frith & Frith, 2010), as well as animacy (Guerrero & Calvillo, 2016). The importance of agency is further substantiated when examining developmental psychology. Infants attend to goal-directed behaviors and are also surprised when goal pursuit is inefficient (Csibra, 2008; Gergely et al., 1995; Sommerville & Woodward, 2005). Agency continues to be important in

the later stages of development. People are intrinsically motivated to be growth-oriented, curious, and focused on their goals (Deci & Ryan, 1985). The belief that one can achieve one's goal contributes positively to human functioning (Holden et al., 1990; Multon et al., 1991; Stajkovic & Luthans, 1998). Therefore, not surprisingly, the agency is related to higher success rates in undertaken activity and indices such as self-esteem, social status, career success, and well-being (Abele & Wojciszke, 2014; Gebauer et al., 2013; Wojciszke et al., 2011).

Given the importance of agency to human functioning, the agency is also signaled in language (see also the spatial cues related to writing direction in the Spatial Agency Bias; Suitner & Maass, 2016; Suitner & Formanowicz, this book), and SGM addresses one of the ways of representing agency in language. Stressing the importance of accumulation and progress of knowledge in science, SGM is a direct descendant of LCM, as verbs reflect the behavior. In contrast, adjectives and nouns pertain to stable dispositional references (Semin & Fiedler, 1991). Building on that notion, the first assumption of SGM is that verbs are the most likely linguistic category to be related to the activity (Formanowicz et al., 2017). Indeed previous research linked verbs to various types of activity, for example, muscular (Foroni & Semin, 2009) or neural activity (Aziz-Zadeh & Damasio, 2008; Cappa & Pulvermüller, 2012; Fischer & Zwaan, 2008; García et al., 2019). Second, we observed that activity is also at the core of agency, as action orientation is present both in definition and in the measurement scales about agency (Abele & Wojciszke, 2007; Abele et al., 2008). By joining these two premises, the SGM states that verbs would be the most likely linguistic category to express agency.

The most direct evidence for SGM comes from studies that ensured the semantic equivalence of stimuli to directly test the link between linguistic categories and agency (Formanowicz et al., 2017; Weis et al., 2022). These studies employed pseudo-words in Polish with a suffix indicating the grammatical category that the words belonged to, e.g., "nefkiczyć" (to nefkit), "nefkickie" (nefkity) and "nefkictwo" (a nefikit). Participants rated words on various dimensions, such as agency, valence, and concreteness. They also assessed communion, a tendency to relate to others - considered a psychological dimension guiding social perception in correspondence with the agency (Abele & Wojciszke, 2007; 2014). Except for agency, none of these dimensions was affected by the grammatical category, indicating that once semantic features are controlled, verbs are no more concrete, communal, or positive than adjectives and nouns but

certainly more agentic. In a study applying real English words, Weis and colleagues (2022) equalized verbs and nouns in terms of frequency of occurrence, concreteness, dominance, arousal, and valence, according to scores available in existing rating norms (Brysbaert et al., 2014; Warriner et al., 2013), and even in terms of the context in which they appear in natural language. Still, verbs were more likely to evoke agentic associations than nouns, confirming that this grammatical category is responsible for a differential agency assignment.

The SGM is not only decoded from language, but references pertaining to the agency are encoded in language using grammatical categories. Specifically, social categories that are typically associated with a high agency, such as ingroup (Formanowicz, 2020), young people (Formanowicz et al., 2017), or men (Caliskan et al., 2022; Formanowicz et al., 2017), are more often expressed with the use of verbs than social categories that are typically linked with a lower agency such as outgroups, the elderly, and women. In these studies, most often verbs are conceptualized as a base form of verbs ("strive" or "help"), given that this form is the most associated with activity and likely also agency (Carrera et al., 2012, 2014). The odds ratio was computed to indicate how likely it was for words representing targets differently associated with agency (such as "man/men" or a "woman/women") to be followed by a verb, considering the base rate of how often these words occurred in the corpora. Across pairs of targets, those more stereotypically agentic were more frequently paired with verbs than those less stereotypically agentic. In an unpublished study testing the validity of all the tools currently used to measure agency in language, verbs outperformed other measures. They were often on par with a newly developed method based on a modern machine learning approach to natural language processing (Nikadon et al., 2025). It is also important to note that the results obtained from natural language correspond with much more fine-grained experimental methodology. Masculine rather than feminine role nouns are expected to serve as thematic agents (connected with a verb) in a sentence (Esaulova et al., 2014; 2017), suggesting a converging pattern of big-scale corpora analyses with eye-tracking studies.

The results mentioned above provide robust evidence for agency being encoded and decoded in language using verbs. People not only reference agency-related phenomena by using verbs but also recognize (and often follow) such agentic references in language. In this way, linguistic agency can be critical in the aforementioned coordinated communication

(see also Echterhoff & Higgins, 2017; Kashima et al., 2018). For example, For example, a recent study found that organizational slogans were more persuasive when using verbs than adjectives (Formanowicz et al., 2021; see also Formanowicz et al., 2024). This can be due to verbs evoking more vivid imagery in correspondence with the dual coding theory (Paivio, 1991). While this has not been investigated about SGM, a study conducted within the LCM framework compared the imaginability of verbs and nouns phrases to observe that the former was more imaginable and thus rated as more trustworthy (Dechêne et al., 2010; Hansen & Wänke, 2010). The verb-agency link not only can be used to study applied effects but also helps explain some of the linguistic underpinnings or correlates of well-established phenomena in social psychology. For example, self is linked to the agency (Abele & Wojciszke, 2007, 2014) because for an individual, the achievement of their own goals and the ability to do that is at the center of attention. Accordingly, it has been found that references to the self were more frequently linked to verbs than references to others. This effect was specific to verbs and not emerging with other linguistic categories (Weis et al., 2022).

Furthermore, ingroup references are linked to verbs more frequently than references to outgroups (Formanowicz et al., 2020), and this effect mirrors other findings in social domains which assign higher agency to an ingroup than outgroup (Fiske et al., 2002; Vaes & Paladino, 2010). This finding can reflect how language contributes to perpetuating intergroup dynamics because agency is related to both positive evaluations (ingroup favoritism) and hierarchical relations, as the correlates of the agency include status and power (Abele & Wojciszke, 2014). Pairing a group with verbs may also signal a real or projected privileged position (Fiske et al., 2002). Overall, there is evidence that verbs can be considered markers of agency, both encoded and decoded by language users.

Agency in a language beyond Social Grammar Model

It is also important to note that the effects under SGM are not the only ones pinpointing the role of agency in language. Agency is, of course, also encoded in semantic references. This is not surprising, however, because many psychological constructs are explicitly mentioned in a given utterance (e.g., "they strive to achieve their goal for the agency"). Such semantic references to the agency are measured through ad hoc created dictionaries comprising agentic traits or phrases (Gaucher et al., 2011;

Nicolas et al., 2021; Pietraszkiewicz et al., 2019), existing dictionaries of Linguistic Inquiry and Word Count (Boyd & Pennebaker, 2015; Tausczik & Pennebaker, 2010), or newly developed measures based on machine-learning (Nikadon et al., 2025). Importantly, these methods capture agency in natural language, albeit with different sensitivity (Nikadon et al., 2025). Still, the use of agentic words has been linked to agency-related factors. An archival analysis of speeches and interviews of prominent physicists, historians, psychologists, and American presidents revealed that controlling for other factors, the use of agentic words (e.g., "achieve" and "strong") was positively related to longevity, most likely due to the relationship between agency, self-regulation, and self-efficacy (Robinson et al., 2016). Agentic language is also used when describing events related to success and effectiveness. For example, upward (vs. downward) trends in the stock market were described with active and agentic (vs. passive and non-agentic) language (Morris et al., 2007).

Similarly, stereotypes referring to groups considered more agentic are captured in natural language, so references to those groups contain more agentic words (e.g., women; Gaucher et al., 2011; professional groups; Pietraszkiewicz et al., 2019; Pietraszkiewicz & Formanowicz, 2023).. Additionally, semantic references to agency correlate with people's actual ratings of agency measured using traditional scales. A recent study compared traditional ratings of professions obtained via questionnaire (Fiske & Dupree, 2014) with how often and how close the profession name appears in a corpus of 800,000 Reuters messages with words typically representing agency (Pietraszkiewicz et al., 2019, Study 3). The correlation for agency reached .55 showing a substantive overlap between the ratings provided by participants on conventional psychological scales and the linguistic representations of the professions in natural language use. Notably, references to the agency are not only encoded but also decoded from language. When seeing words that represent agentic content, "strive" or "act" people tend to engage more in a goal-oriented behavior than when seeing neutral words (Albarracín et al., 2008; Chartier et al., 2020; Weingarten et al., 2016). Thus, similarly to the findings of SGM, references to the semantic agency can be included in our conversations with a high probability that others will read them according to our intention.

It is essential to repeat that many psychological constructs are straightforwardly represented in language, and the agency is not an exception here. What is, however, specific for the agency, is that it has multiple syntactic representations. It is even hypothesized that the adaptive

function of highlighting agents in language indicates the evolutionary underpinnings of the development of syntax (Wilson et al., 2022). On top of being related to the use of verbs, further evidence of syntactic encoding of agency comes from studies on newly emerging sign languages, linguistic topology, and language comprehension, which indicate that there is a basic tendency to rapidly identify the thematic role of agent and patient (for a review see Rissman & Majid, 2019). The tendency for privileged processing of agents in language is a well-documented cross-linguistic phenomenon (Alday et al., 2014; Bornkessel-Schlesewsky & Schlesewsky, 2009), in which the attention paid to agency reflects general attention to cues of agency described earlier. As it is beneficial to quickly establish who is the agent of the sentence to determine who is capable of action, across languages the position of the agent is privileged in sentence processing (Gardelle & Sorlin, 2018).

Agency is also encoded in word order so that when a pair or group of targets is mentioned, those with the higher agency are listed first. This finding has been found either for agentic vs. nonagentic targets such as men listed before women (mainly when they played a dominant role; Kesebir, 2017; see also Peter Hegarty's chapter in this book), but also when cues of agency pertained to close correlates of agency such as social status (McGuire & McGuire, 1992), power (Benor & Levy, 2006), or masculinity (Hegarty et al., 2011). A similar bias was observed in the case of linguistic framing (Bruckmüller & Abele, 2010). There, agentic (or high-status groups) tend to be positioned in a sentence as a standard of comparison. This happens in reference to existing groups such as men and women, with traditional gender hierarchy making men more often the comparison norm than women (Bruckmüller et al., 2012). However, even without references to actual social groups, mere positioning is enough to infer the status and agency of a group. Reading about Maray differ from **Vakuna** vs. Vakuna differ from **Maray** participants were able to determine that a normative group (in bold) has more agency and power.

Conclusion

The numerous layers in which agency is encoded in language reflect the importance assigned to the agency in information processing (Abir et al., 2017; Frith & Frith, 2010) and even language evolution (Wilson et al., 2022). As we need to pay attention to those who could potentially affect our functioning, language prioritizes the processing of relevant signals

(Bornkessel & Schlesewsky, 2006; Muralikrishnan et al., 2015). On a more social level, there can be a privileged transmission of agentic cues in language, assigning a higher weight to such cues in communication, making them relevant when persuading others or engaging them in actions. For example, the defining features of collective action or collective movement are related to people (i.e., a collective) applying their efforts towards a change (i.e., action and movement). Accordingly, the shared sense of identity and agency are among the core predictors of engagement in collective action (Becker & Tausch, 2015; van Zomeren, 2013; van Zomeren, Postmes, et al., 2008; van Zomeren, Spears, et al., 2008). Applying language analysis to collective action research can help understand how an action becomes a collective endeavor, for example, through passionate speeches, provocative and inspiring social media posts, or emotional stories by victims of moral injustice. Indeed, in a yet unpublished set of studies, we have recently conducted, participants whose aim was to encourage others to participate in collective action (vs. to express their thoughts on the importance of such action) used more verbs and other cues of agency in their short texts. Significantly, another group of participants who read these texts in a different study was affected by this language and declared a higher proclivity to act in favor of the environment or as volunteers. We find this important because language is the most accessible and efficient tool for reflecting, transmitting, and shaping socially relevant phenomena (Holtgraves & Kashima, 2008; Anne Maass et al., 2006) and therefore provides an indispensable conceptual and methodological framework for empirical studies of the cognitive processes involved in the formation of collective action or joint action more generally (Sebanz et al., 2006). We consider language to be not only a reflection of reality but also a constituent part of that reality, with speech acts seen as behavioral acts (Searle, 1969). The mere talking about an issue creates its presence in the public sphere, therefore, materializing a thought into action. As such, language practices can contribute to maintaining or transforming the existing social order.

Acknowledgment

References

Abele, A. E., & Wojciszke, B. (2007). Agency and communion from the perspective of self versus others. Journal of Personality and Social Psychology, 93(5), 751–763. https://doi.org/10.1037/0022-3514.93.5.751

Abele, A. E., & Wojciszke, B. (2014). Communal and agentic content in social cognition (Vol. 50, pp. 195–255). Elsevier. https://doi.org/10.1016/B978-0-12-800284-1.00004-7

Abele, A. E., Uchronski, M., Suitner, C., & Wojciszke, B. (2008). Towards an operationalization of the fundamental dimensions of agency and communion: Trait content ratings in five countries considering valence and frequency of word occurrence. European Journal of Social Psychology, 38(7), 1202–1217. https://doi.org/10.1002/ejsp.575

Abir, Y., Sklar, A. Y., Dotsch, R., Todorov, A., & Hassin, R. R. (2017). The determinants of consciousness of human faces. Nature Human Behaviour, 2(3), 194–199. https://doi.org/10.1038/s41562-017-0266-3

Ahearn, L. M. (2001). Language and Agency. Annual Review of Anthropology, 30(1), 109–137. https://doi.org/10.1146/annurev.anthro.30.1.109

Albarracín, D., Handley, I. M., Noguchi, K., McCulloch, K. C., Li, H., Leeper, J., Brown, R. D., Earl, A., & Hart, W. P. (2008). Increasing and decreasing motor and cognitive output: a model of general action and inaction goals. Journal of Personality and Social Psychology, 95(3), 510–523. https://doi.org/10.1037/a0012833

Alday, P. M., Schlesewsky, M., & Bornkessel-Schlesewsky, I. (2014). Towards a computational model of actor-based language comprehension. Neuroinformatics, 12(1), 143–179. https://doi.org/10.1007/s12021-013-9198-x

Assilaméhou, Y., & Testé, B. (2013). How you describe a group shows how biased you are: Language abstraction and inferences about a

speaker's communicative intentions and attitudes toward a group. Journal of Language and Social Psychology, 32(2), 202–211.

Aziz-Zadeh, L., & Damasio, A. (2008). Embodied semantics for actions: findings from functional brain imaging. Journal of Physiology, Paris, 102(1–3), 35–39. https://doi.org/10.1016/j.jphysparis.2008.03.012

Bakan, D. (1966). The duality of human existence: An essay on psychology and religion. Rand Mcnally.

Bandura, A. (2001). Social cognitive theory: an agentic perspective. Annual Review of Psychology, 52, 1–26. https://doi.org/10.1146/annurev.psych.52.1.1

Becker, J. C., & Tausch, N. (2015). A dynamic model of engagement in normative and non-normative collective action: Psychological antecedents, consequences, and barriers. European Review of Social Psychology, 26(1), 43–92. https://doi.org/10.1080/10463283.2015.1094265

Benor, S., & Levy, R. (2006). The chicken or the egg? A probabilistic analysis of english binomials. Language, 82(2), 233–278. https://doi.org/10.1353/lan.2006.0077

Bornkessel, I., & Schlesewsky, M. (2006). The extended argument dependency model: a neurocognitive approach to sentence comprehension across languages. Psychological Review, 113(4), 787–821. https://doi.org/10.1037/0033-295X.113.4.787

Bornkessel-Schlesewsky, I., & Schlesewsky, M. (2009). The Role of Prominence Information in the Real-Time Comprehension of Transitive Constructions: A Cross-Linguistic Approach. Language and Linguistics Compass, 3(1), 19–58. https://doi.org/10.1111/j.1749-818X.2008.00099.x

Boyd, R., & Pennebaker, J. (2015). A way with words: using language for psychological science in the modern era. Consumer Psychology in a Social Media World, 222–236.

Brewer, M. B. (1979). In-group bias in the minimal intergroup situation: A cognitive-motivational analysis. Psychological Bulletin, 86(2), 307–324. https://doi.org/10.1037/0033-2909.86.2.307

Brown, R., & Zagefka, H. (2005). Ingroup affiliations and prejudice. In J. F. Dovidio, P. Glick, & L. A. Rudman (Eds.), On the nature of prejudice (pp. 54–70). Blackwell Publishing Ltd. https://doi.org/10.1002/9780470773963.ch4

Bruckmüller, S., & Abele, A. E. (2010). Comparison focus in intergroup comparisons: who we compare to whom influences who we see as powerful and agentic. Personality and Social Psychology Bulletin,

36(10), 1424–1435. https://doi.org/10.1177/0146167210383581

Bruckmüller, S., Hegarty, P., & Abele, A. E. (2012). Framing gender differences: Linguistic normativity affects perceptions of power and gender stereotypes. European Journal of Social Psychology, 42(2), 210–218. https://doi.org/10.1002/ejsp.858

Bryan, C. J., Walton, G. M., Rogers, T., & Dweck, C. S. (2011). Motivating voter turnout by invoking the self. Proceedings of the National Academy of Sciences of the United States of America, 108(31), 12653–12656. https://doi.org/10.1073/pnas.1103343108

Brysbaert, M., Warriner, A. B., & Kuperman, V. (2014). Concreteness ratings for 40 thousand generally known English word lemmas. Behavior Research Methods, 46(3), 904–911. https://doi.org/10.3758/s13428-013-0403-5

Caliskan, A., Ajay, P. P., Charlesworth, T., Wolfe, R., & Banaji, M. R. (2022). Gender bias in word embeddings: A comprehensive analysis of frequency, syntax, and semantics. Proceedings of the 2022 AAAI/ACM Conference on AI, Ethics, and Society, 156–170. https://doi.org/10.1145/3514094.3534162

Cappa, S. F., & Pulvermüller, F. (2012). Cortex special issue: language and the motor system. Cortex, 48(7), 785–787. https://doi.org/10.1016/j.cortex.2012.04.010

Carnaghi, A., Maass, A., Gresta, S., Bianchi, M., Cadinu, M., & Arcuri, L. (2008). Nomina sunt omina: on the inductive potential of nouns and adjectives in person perception. Journal of Personality and Social Psychology, 94(5), 839–859. https://doi.org/10.1037/0022-3514.94.5.839

Carrera, P., Muñoz, D., Caballero, A., Fernández, I., & Albarracín, D. (2012). The Present Projects Past Behavior into the Future while the Past Projects Attitudes into the Future: How Verb Tense Moderates Predictors of Drinking Intentions. Journal of Experimental Social Psychology, 48(5), 1196–1200. https://doi.org/10.1016/j.jesp.2012.04.001

Carrera, P., Muñoz, D., Caballero, A., Fernández, I., Aguilar, P., & Albarracín, D. (2014). How verb tense affects the construal of action: The simple past tense leads people into an abstract mindset. Psicologica : Revista de Metodologia y Psicologia Experimental, 35(2), 209–223.

Chartier, C. R., Arnal, J. D., Arrow, H., Bloxsom, N. G., Bonfiglio, D. B. V., Brumbaugh, C. C., Corker, K. S., Ebersole, C. R., Garinther, A., Giessner, S. R., Hughes, S., Inzlicht, M., Lin, H., Mercier, B., Metzger, M., Rangel, D., Saunders, B., Schmidt, K., Storage, D., & Tocco, C. (2020). Many labs 5: registered replication of Albarracín et al. (2008),

experiment 5. Advances in Methods and Practices in Psychological Science, 3(3), 332–339. https://doi.org/10.1177/2515245920945963

Csibra, G. (2008). Goal attribution to inanimate agents by 6.5-month-old infants. Cognition, 107(2), 705–717. https://doi.org/10.1016/j.cognition.2007.08.001

Dechêne, A., Stahl, C., Hansen, J., & Wänke, M. (2010). The truth about the truth: A meta-analytic review of the truth effect. Personality and Social Psychology Review, 14(2), 238–257. https://doi.org/10.1177/1088868309352251

Deci, E. L., & Ryan, R. M. (1985). Intrinsic motivation and selfdetermination in human behavior. Plenum.

Dragojevic, M., Sink, A., & Mastro, D. (2017). Evidence of linguistic intergroup bias in U.S. print news coverage of immigration. Journal of Language and Social Psychology, 36(4), 462–472. https://doi.org/10.1177/0261927X16666884

Echterhoff, G., & Higgins, E. T. (2017). Creating shared reality in interpersonal and intergroup communication: the role of epistemic processes and their interplay. European Review of Social Psychology, 28(1), 175–226. https://doi.org/10.1080/10463283.2017.1333315

Esaulova, Y., Reali, C., & von Stockhausen, L. (2014). Influences of grammatical and stereotypical gender during reading: eye movements in pronominal and noun phrase anaphor resolution. Language, Cognition and Neuroscience, 29(7), 781–803. https://doi.org/10.1080/01690965.2013.794295

Esaulova, Y., Reali, C., & Von Stockhausen, L. (2017). Prominence of gender cues in the assignment of thematic roles in German. Applied Psycholinguistics, 38(5), 1133–1172. https://doi.org/10.1017/S014271641700008X

Feldman, L. B., Basnight-Brown, D. M., & Pastizzo, M. J. (2006). Semantic influences on morphological facilitation. Modern Linguistics, 1(1), 59–84. https://doi.org/10.1075/ml.1.1.06fel

Fiedler, K. (2008). Language: A toolbox for sharing and influencing social reality. Perspectives on psychological science, 3(1), 38-47.

Fiedler, K. (2008). The Implicit Meta-Theory That Has Inspired and Restricted LCM Research. Journal of Language and Social Psychology, 27(2), 182–196. https://doi.org/10.1177/0261927X07313656

Fischer, M. H., & Zwaan, R. A. (2008). Embodied language: A review of the role of the motor system in language comprehension. Quarterly Journal of Experimental Psychology, 61(6), 825–850. https://doi.org/10.1080/17470210701623605

Fiske, S. T., Cuddy, A. J. C., Glick, P., & Xu, J. (2002). A model of (often mixed) stereotype content: competence and warmth respectively follow from perceived status and competition. Journal of Personality and Social Psychology, 82(6), 878–902. https://doi.org/10.1037//0022-3514.82.6.878

Fiske, S. T., & Dupree, C. (2014). Gaining trust as well as respect in communicating to motivated audiences about science topics. Proceedings of the National Academy of Sciences of the United States of America, 111 Suppl 4, 13593–13597. https://doi.org/10.1073/pnas.1317505111

Formanowicz, M. (2020). Verb Intergroup Bias: Verbs Are Used More Often in Reference to In-Groups than Out-Groups. Social Psychological and Personality Science, 11(6), 854–864. https://doi.org/10.1177/1948550619893957

Formanowicz, M., Beneda, M., Witkowska, M., Nikadon, J., & Suitner, C. (2024). Mobilize is a verb: the use of verbs and concrete language is associated with authors' and readers' perceptions of a text's action orientation and persuasiveness. Personality and Social Psychology Bulletin. https://doi.org/10.1177/01461672241238418

Formanowicz, M., Roessel, J., Suitner, C., & Maass, A. (2017). Verbs as linguistic markers of agency: The social side of grammar. European Journal of Social Psychology, 47(5), 566–579. https://doi.org/10.1002/ejsp.2231

Formanowicz, M., Pietraszkiewicz, A., Roessel, J., Suitner, C., Witkowska, M., & Maass, A. (2021). "Make it Happen!" Social Psychology, 52(2), 75–89. https://doi.org/10.1027/1864-9335/a000435

Foroni, F., & Semin, G. R. (2009). Language that puts you in touch with your bodily feelings: the multimodal responsiveness of affective expressions. Psychological Science, 20(8), 974–980. https://doi.org/10.1111/j.1467-9280.2009.02400.x

Foster-Hanson, E., Cimpian, A., Leshin, R. A., & Rhodes, M. (2020). Asking children to "be helpers" can backfire after setbacks. Child Development, 91(1), 236–248. https://doi.org/10.1111/cdev.13147

Frith, U., & Frith, C. (2010). The social brain: allowing humans to boldly go where no other species has been. Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences, 365(1537), 165–176. https://doi.org/10.1098/rstb.2009.0160

García, A. M., Moguilner, S., Torquati, K., García-Marco, E., Herrera, E., Muñoz, E., Castillo, E. M., Kleineschay, T., Sedeño, L., & Ibáñez, A. (2019). How meaning unfolds in neural time: Embodied reactivations

can precede multimodal semantic effects during language processing. Neuroimage, 197, 439–449. https://doi.org/10.1016/j. neuroimage.2019.05.002

Gardelle, L., & Sorlin, S. (2018). Introduction. International Journal of Language and Culture, 5(2), 133–162. https://doi.org/10.1075/ ijolc.00004.gar

Gaucher, D., Friesen, J., & Kay, A. C. (2011). Evidence that gendered wording in job advertisements exists and sustains gender inequality. Journal of Personality and Social Psychology, 101(1), 109–128. https:// doi.org/10.1037/a0022530

Gebauer, J. E., Wagner, J., Sedikides, C., & Neberich, W. (2013). Agency-communion and self-esteem relations are moderated by culture, religiosity, age, and sex: evidence for the "self-centrality breeds self-enhancement" principle. Journal of Personality, 81(3), 261–275. https://doi.org/10.1111/j.1467-6494.2012.00807.x

Gelman, S. A., & Heyman, G. D. (1999). Carrot-Eaters and Creature-Believers: The Effects of Lexicalization on Children's Inferences About Social Categories. Psychological Science, 10(6), 489–493. https://doi. org/10.1111/1467-9280.00194

Gerber, A., Huber, G., & Fang, A. (2017). Do subtle linguistic interventions priming a social identity as a voter have outsized effects on voter turnout? evidence from a new replication experiment. Political Psychology, 39(4), 925–938. https://doi.org/10.1111/pops.12446

Gerber, A. S., Huber, G. A., Biggers, D. R., & Hendry, D. J. (2016). A field experiment shows that subtle linguistic cues might not affect voter behavior. Proceedings of the National Academy of Sciences of the United States of America, 113(26), 7112–7117. https://doi.org/10.1073/ pnas.1513727113

Gergely, G., Nádasdy, Z., Csibra, G., & Bíró, S. (1995). Taking the intentional stance at 12 months of age. Cognition, 56(2), 165–193. https://doi.org/10.1016/0010-0277(95)00661-h

Gorham, B. W. (2006). News Media's Relationship With Stereotyping: The Linguistic Intergroup Bias in Response to Crime News. Journal of Communication, 56(2), 289–308. https://doi.org/10.1111/j.1460-2466.2006.00020.x

Guerrero, G., & Calvillo, D. P. (2016). Animacy increases second target reporting in a rapid serial visual presentation task. Psychonomic Bulletin & Review, 23(6), 1832–1838. https://doi.org/10.3758/s13423-016-1040-7

Hansen, J., & Wänke, M. (2010). Truth from language and truth from

fit: the impact of linguistic concreteness and level of construal on subjective truth. Personality and Social Psychology Bulletin, 36(11), 1576–1588. https://doi.org/10.1177/0146167210386238

Hegarty, P., Watson, N., Fletcher, L., & McQueen, G. (2011). When gentlemen are first and ladies are last: effects of gender stereotypes on the order of romantic partners' names. The British Journal of Social Psychology / the British Psychological Society, 50(Pt 1), 21–35. https://doi.org/10.1348/014466610X486347

Holden, G., Moncher, M. S., Schinke, S. P., & Barker, K. M. (1990). Self-efficacy of children and adolescents: a meta-analysis. Psychological Reports, 66(3 Pt 1), 1044–1046. https://doi.org/10.2466/pr0.1990.66.3.1044

Holtgraves, T. M., & Kashima, Y. (2008). Language, meaning, and social cognition. Personality and Social Psychology Review, 12(1), 73–94. https://doi.org/10.1177/1088868307309605

Kashima, Y., Bratanova, B., & Peters, K. (2018). Social transmission and shared reality in cultural dynamics. Current Opinion in Psychology, 23, 15–19. https://doi.org/10.1016/j.copsyc.2017.10.004

Kesebir, S. (2017). Word order denotes relevance differences: The case of conjoined phrases with lexical gender. Journal of Personality and Social Psychology, 113(2), 262–279. https://doi.org/10.1037/pspi0000094

Kousta, S.-T., Vigliocco, G., Vinson, D. P., Andrews, M., & Del Campo, E. (2011). The representation of abstract words: Why emotion matters. Journal of Experimental Psychology: General, 140(1), 14–34. https://doi.org/10.1037/a0021446

Maass, A., Arcuri, L., & Suitner, C. (2014). Shaping intergroup relations through language. In T. M. Holtgraves (Ed.), The Oxford handbook of language and social psychology (pp. 157–176). Oxford University Press.

Maass, A., Ceccarelli, R., & Rudin, S. (1996b). Linguistic intergroup bias: Evidence for in-group-protective motivation. Journal of Personality and Social Psychology, 71(3), 512–526. https://doi.org/10.1037/0022-3514.71.3.512

Maass, A., Karasawa, M., Politi, F., & Suga, S. (2006). Do verbs and adjectives play different roles in different cultures? A cross-linguistic analysis of person representation. Journal of Personality and Social Psychology, 90(5), 734–750. https://doi.org/10.1037/0022-3514.90.5.734

Maass, A., Milesi, A., Zabbini, S., & Stahlberg, D. (1995). Linguistic intergroup bias: differential expectancies or in-group protection? Journal of Personality and Social Psychology, 68(1), 116–126. https://

doi.org/10.1037/0022-3514.68.1.116

Maass, A., Salvi, D., Arcuri, L., & Semin, G. (1989). Language use in intergroup contexts: the linguistic intergroup bias. Journal of Personality and Social Psychology, 57(6), 981–993. https://doi.org/10.1037/0022-3514.57.6.981

McGuire, W. J., & McGuire, C. V. (1992). Cognitive-versus-affective positivity asymmetries in thought systems. European Journal of Social Psychology, 22(6), 571–591.

Morris, M. W., Sheldon, O. J., Ames, D. R., & Young, M. J. (2007). Metaphors and the market: Consequences and preconditions of agent and object metaphors in stock market commentary. Organizational Behavior and Human Decision Processes, 102(2), 174–192. https://doi.org/10.1016/j.obhdp.2006.03.001

Multon, K. D., Brown, S. D., & Lent, R. W. (1991). Relation of self-efficacy beliefs to academic outcomes: A meta-analytic investigation. Journal of Counseling Psychology, 38(1), 30–38. https://doi.org/10.1037/0022-0167.38.1.30

Muralikrishnan, R., Schlesewsky, M., & Bornkessel-Schlesewsky, I. (2015). Animacy-based predictions in language comprehension are robust: contextual cues modulate but do not nullify them. Brain Research, 1608, 108–137. https://doi.org/10.1016/j.brainres.2014.11.046

New, J., Cosmides, L., & Tooby, J. (2007). Category-specific attention for animals reflects ancestral priorities, not expertise. Proceedings of the National Academy of Sciences of the United States of America, 104(42), 16598–16603. https://doi.org/10.1073/pnas.0703913104

Nicolas, G., Bai, X., & Fiske, S. T. (2021). Comprehensive stereotype content dictionaries using a semi-automated method. European Journal of Social Psychology, 51(1), 178–196. https://doi.org/10.1002/ejsp.2724

Nikadon, J., Suitner, C., Erseghe, T., Dzanko, L., & Formanowicz, M. (2025). BERTAgent: The Development of a Novel Tool to Quantify Agency in Textual Data. Journal of Experimental Psychology: General. https://doi.org/10.1037/xge0001740

Oppenheimer, D. M., & Frank, M. C. (2008). A rose in any other font would not smell as sweet: effects of perceptual fluency on categorization. Cognition, 106(3), 1178–1194. https://doi.org/10.1016/j.cognition.2007.05.010

Paivio, A. (1991). Dual coding theory: Retrospect and current status. Canadian Journal of Psychology/Revue Canadienne de Psychologie, 45(3), 255–287. https://doi.org/10.1037/h0084295

Pietraszkiewicz, A., & Formanowicz, M. (2023). Verbs are associated with agency. Social Psychology, 54(5), 271–282. https://doi.org/10.1027/1864-9335/a000522

Pietraszkiewicz, A., Formanowicz, M., Gustafsson Sendén, M., Boyd, R. L., Sikström, S., & Sczesny, S. (2019). The big two dictionaries: Capturing agency and communion in natural language. European Journal of Social Psychology, 49(5), 871–887. https://doi.org/10.1002/ejsp.2561

Rhodes, M., Leslie, S.-J., Yee, K. M., & Saunders, K. (2019). Subtle linguistic cues increase girls' engagement in science. Psychological Science, 30(3), 455–466. https://doi.org/10.1177/0956797618823670

Rissman, L., & Majid, A. (2019). Thematic roles: Core knowledge or linguistic construct? Psychonomic Bulletin & Review, 26(6), 1850–1869. https://doi.org/10.3758/s13423-019-01634-5

Robinson, M. D., Bair, J. L., Persich, M. R., & Moen, N. R. (2016). Linguistic Agency and Life-Span Longevity. Psychosomatic Medicine, 78(7), 829–834. https://doi.org/10.1097/PSY.0000000000000337

Ruscher, J. B., & Tipler, C. N. (2018). Linguistic intergroup bias about the 2016 U.S. presidential candidates as a function of political ideology. Analyses of Social Issues and Public Policy, 18(1), 61–80. https://doi.org/10.1111/asap.12149

Schwanenflugel, P. J., & Stowe, R. W. (1989). Context availability and the processing of abstract and concrete words in sentences. Reading Research Quarterly, 24(1), 114. https://doi.org/10.2307/748013

Searle, J. R. (1969). Speech Acts: An essay in The philosophy of Language. Cambridge University Press. https://doi.org/10.1017/CBO9781139173438

Sebanz, N., Bekkering, H., & Knoblich, G. (2006). Joint action: bodies and minds moving together. Trends in Cognitive Sciences, 10(2), 70–76. https://doi.org/10.1016/j.tics.2005.12.009

Semin, Gün R. (2000). Agenda 2000? communication: language as an implementational device for cognition. European Journal of Social Psychology, 30(5), 595–612. https://doi.org/10.1002/1099-0992(200009/10)30:5<595::AID-EJSP23>3.0.CO;2-A

Semin, Gün R. (2008). Language Puzzles. Journal of Language and Social Psychology, 27(2), 197–209. https://doi.org/10.1177/0261927X07313664

Semin, Gün R., & Cacioppo, J. T. (2008). Grounding Social Cognition: Synchronization, Coordination, and Co-Regulation. In Embodied grounding: social, cognitive, affective, and neuroscientific approaches (pp. 119–147). Cambridge University Press. https://doi.org/10.1017/CBO9780511805837.006

Semin, Gün R., & Fiedler, K. (1988). The cognitive functions of linguistic categories in describing persons: Social cognition and language. Journal of Personality and Social Psychology, 54(4), 558–568. https://doi.org/10.1037/0022-3514.54.4.558

Semin, Gün R., & Fiedler, K. (1991). The linguistic category model, its bases, applications and range. European Review of Social Psychology, 2(1), 1–30. https://doi.org/10.1080/14792779143000006

Simion, F., Regolin, L., & Bulf, H. (2008). A predisposition for biological motion in the newborn baby. Proceedings of the National Academy of Sciences of the United States of America, 105(2), 809–813. https://doi.org/10.1073/pnas.0707021105

Sommerville, J. A., & Woodward, A. L. (2005). Pulling out the intentional structure of action: the relation between action processing and action production in infancy. Cognition, 95(1), 1–30. https://doi.org/10.1016/j.cognition.2003.12.004

Stajkovic, A. D., & Luthans, F. (1998). Self-efficacy and work-related performance: A meta-analysis. Psychological Bulletin, 124(2), 240–261. https://doi.org/10.1037/0033-2909.124.2.240

Suitner, C., & Maass, A. (2016). Spatial Agency Bias (Vol. 53, pp. 245–301). Elsevier. https://doi.org/10.1016/bs.aesp.2015.09.004

Tajfel, H., & Turner, J. (1979). An integrative theory of inter-group conflict. In W. Austin & S. Worchel (Eds.), The social psychology of intergroup relations. Brooks/Cole.

Tausczik, Y. R., & Pennebaker, J. W. (2010). The psychological meaning of words: LIWC and computerized text analysis methods. Journal of Language and Social Psychology, 29(1), 24–54. https://doi.org/10.1177/0261927X09351676

Vaes, J., & Paladino, M. P. (2010). The uniquely human content of stereotypes. Group Processes & Intergroup Relations, 13(1), 23–39. https://doi.org/10.1177/1368430209347331

Vainapel, S., Shamir, O. Y., Tenenbaum, Y., & Gilam, G. (2015). The dark side of gendered language: The masculine-generic form as a cause for self-report bias. Psychological Assessment, 27(4), 1513–1519. https://doi.org/10.1037/pas0000156

van Zomeren, M., Postmes, T., & Spears, R. (2008). Toward an integrative social identity model of collective action: a quantitative research synthesis of three socio-psychological perspectives. Psychological Bulletin, 134(4), 504–535. https://doi.org/10.1037/0033-2909.134.4.504

van Zomeren, M. (2013). Four Core Social-Psychological Motivations to Undertake Collective Action. Social and Personality Psychology

Compass, 7(6), 378–388. https://doi.org/10.1111/spc3.12031

van Zomeren, M., Spears, R., & Leach, C. W. (2008). Exploring psychological mechanisms of collective action: does relevance of group identity influence how people cope with collective disadvantage? The British Journal of Social Psychology / the British Psychological Society, 47(Pt 2), 353–372. https://doi.org/10.1348/014466607X231091

Walton, G. M., & Banaji, M. R. (2004). Being what you say: the effect of essentialist linguistic labels on preferences. Social Cognition, 22(2), 193–213. https://doi.org/10.1521/soco.22.2.193.35463

Warriner, A. B., Kuperman, V., & Brysbaert, M. (2013). Norms of valence, arousal, and dominance for 13,915 English lemmas. Behavior Research Methods, 45(4), 1191–1207. https://doi.org/10.3758/s13428-012-0314-x

Weingarten, E., Chen, Q., McAdams, M., Yi, J., Hepler, J., & Albarracín, D. (2016). From primed concepts to action: A meta-analysis of the behavioral effects of incidentally presented words. Psychological Bulletin, 142(5), 472–497. https://doi.org/10.1037/bul0000030

Weis, P. P., Nikadon, J., Herbert, C., & Formanowicz, M. (2022). The verb–self link: An implicit association test study. Psychonomic Bulletin & Review, 29(5), 1946–1959. https://doi.org/10.3758/s13423-022-02105-0

Wigboldus, D. H. J., Semin, G. R., & Spears, R. (2000). How do we communicate stereotypes? Linguistic bases and inferential consequences. Journal of Personality and Social Psychology, 78(1), 5.

Wilson, V. A. D., Zuberbühler, K., & Bickel, B. (2022). The evolutionary origins of syntax: Event cognition in nonhuman primates. Science Advances, 8(25), eabn8464. https://doi.org/10.1126/sciadv.abn8464

Witkowska, M., Dołżycka, J., Suitner, C., & Formanowicz, M. (2024). The Grammar of Persuasion: A Meta-Analytic Review Disconfirming the Role of Nouns as Linguistic Cues of Subsequent Behavior. Journal of Language and Social Psychology. https://doi.org/10.1177/0261927X241234845

Wojciszke, B., Baryla, W., Parzuchowski, M., Szymkow, A., & Abele, A. E. (2011). Self-esteem is dominated by agentic over communal information. European Journal of Social Psychology, 41(5), 617–627. https://doi.org/10.1002/ejsp.791

8. Let's Make Some Order—The Role of (W)ordering in Social Cognition

Maria Laura Bettinsoli
University of Padova, Italy

'Words are sacred. They deserve respect. If you get the right ones, in the right order, you can nudge the world a little.'

Tom Stoppard

Imagine you need a cup placed on the shelf close to your friend. You can ask your friend "give me the cup, please" but also "the cup, give it to me, please". Both cases, you will eventually get the cup, but the arrangement of words may correspond to a different level of effectiveness in communicating your request. In fact, languages are claimed to be shaped for efficient use to allow speakers balancing their needs to both minimize production effort and -at the same time- be understood by their interlocutors (e.g., Givón, 1991; Gibson et al., 2019; Haspelmath, 2021).

Social psychology and psycholinguistics have found accumulating evidence that language is a powerful tool not only for communicating, constructing, and representing meaning, but also for transforming social reality. One facet of language is likely seldom thought of - the order in which information comes to our attention. It is one linguistic tool to establish temporal and causal links between elements, but it may also strategically communicate the importance attributed to any single element. For instance, we can mention one object or person before the other prioritizing the most important to us. Put simply, order appears to contribute to both guiding attention and interpreting reality. Language users should be sensitive to the incremental nature of language processing, as it determines the order in which information becomes available to the

listeners, influencing the way in which they perceive speakers' narration. Many studies showed that cognitive processes are indeed influenced by some forms of order. For instance, various studies on the relationship between language scripts – the direction in which a language is written and read- and spatial biases demonstrate that there is an order according to which we interpret and represent social actions (e.g., Maass, Pagani & Berta, 2007; Maass, Suitner, Nadhmi, 2014), imagine the time line (e.g., Boroditsky, 2001), explore space (e.g., Maass, Suitner and De-conchy, 2014), and position our ingroup (e.g., Bettinsoli et al., 2022). For instance, Italian speakers recall better objects positioned on the left (vs. right), whereas Arabic speakers reveal the opposite pattern (Bettinsoli et al., 2019). These differences in social representation, space and time conceptualization, and objects recall can find the most plausible of explanations in research on the Spatial Agency Bias (i.e., SAB, see Chapter 9 by Suitner & Formanowicz in this book for a more detailed review), which suggests that the way in which people, for instance, pay attention to and memorize objects' position follows the order of their languages' script trajectory (i.e., horizontally from left to right for Italian vs. from right to left for Arabic). For instance, Italian (vs. Arabic) speakers recall more accurately objects placed on the left (vs. right) of a grid, presumably because they will start memorizing objects -thus paying their attention- from the left (Italian) or from the right (Arabic) side, respectively (e.g., Bettinsoli et al., 2019). We thus may infer a mutual influence between temporal and spatial dimensions and order both at concrete and abstract levels—that is, on the one hand order seems to imply temporal and spatial aspect, on the other hand both time and space elicit the idea of order. More specifically, temporal (e.g., before/after) and spatial (e.g., left/right) cues depend on a definite order, however, by reading for instance "father and son", speakers may extrapolate information that a) are related to temporal succession (i.e., father was born before the son) and b) speak for a higher level of authority and agency conveyed to fathers as compared to sons, thus, fathers could be mentioned first to signal a hierarchical order. For this reason, several linguistic binomials (e.g., 'men and women') are believed to be syntactically frozen because they reflect a hierarchical semantic order (i.e., men are more powerful than women; Mollin, 2012).

  After briefly defining word order and its variations across languages, I will focus on recent research showing how the order in which words are placed- structurally or strategically- in a sentence contributes to meaning creation and interpretation. This chapter aims at illustrating that such

an intrinsic and subtle feature of language—which is typically taken for granted and seldom reflected—plays a critical role in influencing social cognitive processes.

Word Order: definition, distributions, and variations

Although subtle, word order is a relevant tool influencing the way we create and process language, and this occurs since early age. Infants, indeed, process basic word order very early in development—as they tend to follow the word order rules of a language from their first multi-word utterances (Brown, 1973; Gómez & Gerken, 2000).

Word Order is a typological property of languages, which refers to the combination of the three basic elements Subject (S), Object (O) and Verb (V) in six possible logical orders: SOV, SVO, VSO, VOS, OVS, and OSV (Dryer, 2011). On one level, the combination of the three basic elements represents a mere grammatical and syntactic aspect, namely, we create sentences that follow syntax rules. On a second level both when (i.e., first or last) and where (i.e., right or left) these elements are placed refers to the higher-level cognitive processes of meaning creation and meaning comprehension. For instance, the way in which linguistic binomials are interpreted (i.e., "father and son"; e.g., Hegarty et al., 2011; Mollin, 2012) seems to reflect the order in which elements are positioned and presented in a sentence. The first element mentioned corresponds to the first element positioned and together temporal and spatial cues contribute to determine the importance one may attribute to it (e.g., Bettinsoli et al., 2015).

Several cross-linguistic studies have demonstrated that the distribution of word orders across the world languages is not equal and regular (Dryer, 2011; Greenberg, 1963; Tomlin, 1986), and they have demonstrated a consistent preference for SOV (e.g., 'Mary cats loves") and SVO (e.g., "Mary loves cats") orders in arranging the sentences' syntactic elements (Goldin-Meadow et al., 2008; Goldin-Meadow & Alibali, 2013; Gibson et al., 2013; Hall, Mayberry, & Ferreira, 2013; Langus & Nespor, 2010). There are few approaches that tried to explain the prevalence of SOV and SVO over the other word orders within the field of linguistics. Although they do not necessarily always exclude each other, the explanations provided from these approaches space from the existence of innate universal grammar (i.e., generativist approach; e.g., Chomsky, 1986; Gibson, Piantadosi, Brink, Bergen, Lim and Saxe, 2013) to the functional role that some orders

serve better than others based on identified principles, which tie word order to meaning creation and interpretation (i.e., functionalist approach; Song, 2000; Tomlin, 1986). Accordingly, the ordering of words reflects its function when (a) old information comes before new one (Theme-first principle); (b) subjects precede objects (Animated-first principle, Tomlin 1986), and (c) objects tend to be more closely tied to verbs than subjects (i.e., Verb-Object bonding, which is supported by phenomena such as object incorporation being found in many languages) to better distinguish the recipient (i.e., object) from the agent (i.e., subject) of an action and reduce ambiguity. Since these three principles are only satisfied in SOV and SVO (and not in all the other four possible combinations of order), these orders can be considered as the most functional for communicative purposes. Yet, an unequal distribution of orders might not only be a matter of following some identified principles, but it must also consider that language is strictly connected to the complexity of representations in human mind (i.e., connectivist approach; e.g., Tabullo et al., 2012). Accordingly, the prevalence of some orders over the others can also be the result of different levels of order complexity. A series of studies (e.g., Tabullo et al., 2012; Tily et al., 2011) supports the idea that frequency distribution of SOV and SVO is determined by learnability—that is, because some orders (e.g., SVO) more than others (e.g., VOS) match information processing, functional principles, and reduce ambiguity in communication, they are more natural and easier to learn (Grüning, 2003). Consequently, verb initial word orders might be less frequent presumably because they are more difficult to learn and less efficient in disambiguating communication (Grüning, 2003; Tabullo et al., 2012; however, see Lupyan & Christiansen, 2002; Tily, Frank, & Jaeger, 2011 for contradicting evidence).

Intriguingly, some studies showed that a preference for SOV order is related to improvised gestured communication (Goldin- Meadow et al., 2008; Langus and Nespor, 2010). Sign languages spontaneously emerging within deaf populations have shown a consistent preference for SOV, regardless the order of spoken languages in the environment (Sandler et al., 2005; Goldin-Meadow et al., 2008; Goldin-Meadow & Alibali, 2013). This is especially true when only one sentence is involved, whereas when gesturing events involved more complex structures (e.g., more than one sentence), participants are more likely to follow SVO rather SOV order (Langus and Nespor, 2010). The authors conducted a series of experiments with Italian (SVO) and Turkish (SOV) participants, which included gesture production. Interestingly, in gesturing complex events involving

more than one sentence, the typical construction of SOV languages was ignored: when subordinate clauses were embedded within the main one, participants gestured the subordinate clause immediately after the main one, following SVO rather than SOV order. The cross-linguistic robustness of these results suggests a preference for different orders that can not be attributed to the participant's native languages, given that Italian is a SVO and Turkish a SOV language. Why is that the case? Langus and Nespor (2010) agree with Goldin-Meadow and colleagues (2008) in explaining the preference for SOV in improvised gestured communication: the sequence of subjects, objects and verbs seems to match the order with which events are mentally represented. In line with these findings, Gibson and colleagues (2013) found that English (SVO), Chinese (SVO) and Korean speakers (SOV) -asked to gesture events- showed a cross-linguistically consistent preference for the SOV order when subject and the object in the events were inanimate or when the agent was animate and the patient inanimate (e.g., "The storm the roof damaged" or "John the chair moved"). In contrast, when both the subject and the object were human, participants were more likely to use the SVO word order, especially in events where it is important to distinguish who does what to whom (e.g., "Peter pushes Erick"). It might be the case that SVO helped participants to reduce 'noise'—that is, in SOV (e.g., Peter Erick pushes) two potential agents (i.e., who does the action) are presented at the same side of the verb and this might increase ambiguity, which intuitively appears very important to solve when humans are involved. Additional evidence for this mechanism comes from a study (Hall et al., 2013), which confirmed the SVO preference in describing semantically reversible events, namely when a verb can equally refer to the subject or the object (i.e., "Peter pushes Erik", or "Erik pushes Peter"). Authors found that, when asked to gesture transitive (vs. intransitive) action, individuals were more likely to put themselves in the agent role to avoid the proximity of subjects and objects determined by SOV order, which could generate ambiguity regarding role attribution (i.e., role conflict). In fact, if in intransitive events the agent of an action is always clear (e.g., "John reads the book" where the book cannot read John), in transitive events it is not always the case (e.g., "John calls Chris", where the action might equally be performed by either John or Chris). Therefore, speakers may prefer SVO over SOV order to reduce noise and distinguish the subject from an object of an action (Hall et al., 2013). In sum, literature seems to suggest that the reason why SVO and SOV are more common than other types of order might lay

in the fact that these two orders better satisfy functional principles and information processing, and they are particularly efficient in reducing ambiguity in communication.

Importantly, languages vary (also) with respect to the amount and type of word order deviations they tolerate (i.e., word order flexibility; e.g., Song, 2001), with some languages being relatively rigid and others allowing speakers to deviate from the grammatically correct canonical word order to transmit messages in a more pragmatic – or strategic- way. To get back to the initial example, for instance, we may ask a friend to give us a cup, either following a canonical order (i.e., "[you] give me the cup, please") or pointing the attention strategically on what you need first (i.e., "the cup, [you] give it to me please"). Although the second (vs. first) option is not grammatically correct in many languages, it might serve the purpose of your request better in firstly pointing your interlocutor attention on what you need. Many languages have, indeed, structures which are highly creative, but also extremely infrequent. For instance, the sentence 'On the table is the pen' involves locative inversion, in which the positions of the subject 'the pen' and the locative phrase 'on the table' are switched relative to canonical English word order, and although this construction is rare, it is fully understandable by English speakers. Notably, English has a quite rigid word order because variations might change the meaning of the sentence: when the order of transitive sentences is changed (e.g., "Mary hits Jack"   "Jack hits Mary"), the direction of the action also changes (Kaiser & Trueswell, 2004). In fact, since in English gender, numerical, and case marking are absent, word order helps to encode how elements are related revealing the direction and meaning of a sentence. Thus, the incremental comprehension relies on word order to predict 'who did what to whom' (e.g., Lamers and De Swart, 2012; see Bader and Bayer, 2006 for a review). Recently, a study (Suitner et al., 2021) showed that even the intensity of spatial biases is influenced by word order flexibility—that is, for instance, spatially representing the subject of an action to the left of the object might be more frequent in those left-to-right languages that allow a lower (i.e., Italian; vs. higher: English) degree of order flexibility. Specifically, a cross-linguistic study on 14 European languages, which vary in word order flexibility, revealed that the rightward bias in drawings of interactions between two people (agent and recipient) was weaker in more flexible languages.

Several works have tried to explain the psycholinguistic principles that drive comprehension of alternative word orders different from the

canonical of a specific language. Experimental evidence suggests that alternative word orders are more difficult to process than canonical ones, as reflected by longer reading times, response times and lower accuracy rates on different tasks and this might be explained by ease of processing due to the frequency of usage of a language (e.g., Bader and Meng, 1999; Gattei et al., 2017). This holds similarly also across different languages with different degrees of flexibility (e.g., Italian and German; Dröge et al., 2014; Bornkessel et al., 2005). When alternative forms are more difficult to process then? Notably, Bates and colleagues (1982) showed that some languages allow the use of different orders when it is possible to rely on alternative features of language (e.g., gender and numerical agreement). They studied sentence comprehension with Italian, English, and German speakers and found that English participants interpreted simple sentences mainly by relying on word order, Italian participants made use of semantic cues (i.e., relying on the words' meaning), and German speakers relied on case marking (i.e., inflections of the words). According to the authors, speakers of languages that allow greater word order flexibility might use semantic/pragmatic cues in language creation and comprehension without needing to follow the canonical order (Bates et al., 1982).

Generally, even languages with a predominant word order allow exceptions, especially in spoken language and in languages with case marking (Bentz & Christiansen, 2010). Rather than arranging words in line with correct syntax (canonical order), speakers may use a different order to draw attention to a specific element of the sentence (Johnson & Braber, 1998; Song, 2001. Thus, one of the main challenges in studying word ordering is to address the (mutual) contribution of both grammatical and communicative functions in interpreting and creating meaning. In the following section, I will review previous research addressing the relation between word order and (social) cognitive processes.

Word order and (social) cognitive processes—a mutual influence in meaning creation and interpretation

Can word order affect how information is created or interpreted? Is there a difference whether you read "the students the teacher criticizes" or "the teacher criticizes the students"? This was tested in two cross-linguistic (Italian and English) studies by Bettinsoli and colleagues (2015), where a translations paradigm allowed to manipulate all the possible six orders in which S, O, and V can appear in transitive active sentences

(e.g., "The mom has taken the two children to the concert") while ignoring additional elements (e.g., adverbial modifiers) or secondary clauses. Basically, Italian and English participants with no knowledge of Chinese language were provided with six sentences and told that they were translated literally from Chinese. Participants were presented with the 'original' Chinese sentence above and the translation in their own language below. In this way, authors manipulated the order in which the elements might appear violating the canonical orders (e.g., "Has taken to the concert the two children the mom"). Authors' primary interest was measuring the basic attribution of responsibility for the situation described in the sentences. They focused on the three basic attributions of responsibility to the agent (S in active sentences; e.g., "How much do you think the action is due to [the mom]"), the patient (O in active sentences; e.g., "How much do you think the action is due to [the children]") and the situation in which the action took place via action verbs (V; e.g., "How much do you think the action is a reaction due to the situation [has taken to the concert]?"), as verbs tend to communicate the meaning of the situation (Semin & Fiedler, 1988). The findings suggest, indeed, that the order in which S, O, and V are arranged influences people's interpretation of the described events, directly affecting the causal inferences people draw about everyday events. Specifically, a given element was more likely perceived as the cause of the event when it occurred in the first rather than in a later (2nd or 3rd) position. For instance, when the children (O) appeared in the first position (i.e., "the children the mom has taken to the concert" or "the children has taken to the concert the mom"), they were attributed more responsibility for being taken to the concert than when they appeared in a later position (e.g., "the mom has taken to the concert the children"). This does not mean that patient (O) or situational (V) attributions override higher responsibility attributions to the agent (S) only because the object or verb appears in the first position—that is, the subject has always been identified as the most responsible for the described situations. In fact, in all order combinations of the study, agent (S) attributions were higher than patient (O) or situational (V) attributions. However, when the object or the verbs were moved in first position, they were attributed more responsibility as compared to when they were placed in 2nd or 3rd position. Accordingly, this might also offer a possible explanation for related phenomena such as the tendency to blame the victims (more than the agents of the crime) when the events are described using the passive form of the verbs (e.g., a woman was raped by a stranger).

In fact, when passive forms of the verb are used, the recipient (O) of a situation is syntactically placed in first position. This has been shown to be a systematic strategy for also blaming women in rape events (Bohner, 2001)—that is, placing women in first position increases the likelihood of being perceived as co-responsible for the rape event.

Further empirical evidence for the role of word order in creating meaning comes from several lines of research not only and not directly involving the sequencing of S, O and V, but that speak to the cognitive implications of ordering at a more general level. For instance, research on situation models in text comprehension focuses on clauses and sentences rather than on single words (Johnson-Laird, 1983; Zwaan & Radvansky, 1998). Situation models refer to the mental representations created by the process through which listeners/readers imagine people and situations, position them in time and space, and construct temporal and causal representations of them (Gernsbacher, Goldsmith, & Robertson, 1992; Segal, 1995). When no other information is provided, the temporal and causal relations are established based on the order in which events are presented (iconicity assumption), assigning former (vs. latter) elements temporal and causal primacy—that is, whatever comes first is the cause of what comes later. Conceptually similar findings come from a long tradition of research focusing on the relevance of the first element as driving and influencing the interpretation of subsequent information. For instance, the first information you receive about a target person will be the one influencing the general impression you might form about this person, as you will use it to interpret all the other information (Asch, 1946; Hamilton & Sherman, 1996). The first element encountered when reading or listening attracts greater attention (MacWhinney, 1977), is better remembered (primacy effect) and serves as the starting point for perspective taking and interpretation.

Besides guiding interpretation, word order is relevant in meaning attribution - for instance, in terms of agency and causality- as shown by research on binomial phrases (Hegarty, Mollin, & Foels, 2016; McGuire & McGuire, 1982; Mollin, 2012). Why is it more likely to hear 'fish and chips' and not 'chips and fish'? Early studies on binomial order speak about semantic and phonological constraints that would tend to "freeze" some binomial in a specific order—for instance, in this case by placing an animate object (fish) before an inanimate one (chips) (Cooper and Ross, 1975; McDonald, Bock, & Kelly, 1993). More recent studies (e.g., Mollin, 2012; Hegarty, 2015) show that many binomials have a predom-

inant order, and their degree of reversibility is low because of specific semantic constraints that are linked (also) to society —that is, the first term is the more powerful and agentic relative to hierarchical relations in the real world (e.g., brother and sister, men and women, father and son). For instance, the first mentioned element is perceived as more active and potent (Johnson, 1967), masculine individuals are named (and graphed) before feminine ones (Hegarty & Lemieux, 2011; Hegarty, 2015), higher status groups tend to be mentioned before low status ones (McGuire & McGuire, 1982), and positive nouns precede negative ones (Rozin, Berman, & Royzman, 2010).

It goes without saying that the order in which binomial phrases are constructed -either due to imposed constraints or deliberate strategy - becomes an implicit way to compare two entities. The point of reference will be placed in first position systematically, thus, contributing to different attributional processes (e.g., Pratto, Hegarty, & Korchmaros, 2007; Bruckmüller, Hegarty, & Abele, 2012). Thus, both linguistic and social-psychological work concur in arguing that word order may not be random. Rather, the first element conveys, among others, a relative advantage in terms of agency, power, status, and masculinity. Notably, the archival research by Mollin (2012) has shown that the first element of binomials tends to be chronologically antecedent to the second (e.g., before and after) and the cause rather than the effect (e.g., trial and error). Although temporal ordering does not necessarily imply causal relations, it is a precondition for causal reasoning (i.e., cause before effect) and this may explain why temporal and causal inferences often go hand in hand. Causal relations and inferences are part of people's daily reasoning: we spend much time wondering about causal relations between events, reasons, and reactions in others' behavior, causes and effects of natural catastrophes or physical diseases. Accordingly, it appears natural that human minds are wired to exploit even information such as word order.

To scrutinize these links with causal inferences, it is relevant to distinguish which types of reasoning people might incur while exploring causal relations between event. Tversky and Kahneman (1974) demonstrated that people use their knowledge and experience to make both predictive and diagnostic inferences—that is, one may reason in a predictive way while looking for the likelihood of effects starting from a given cause (i.e., follow a cause-effect order), whereas the diagnostic way is applied when causes of an event are inferred starting from the observed effect (i.e., an effect-cause order). For instance, the predictive reasoning

refers to a situation in which people may look for the possible effects of a certain situation (e.g., "given that they drink alcohol, it is likely their memory is less effective"). In the diagnostic case, by contrast, people may look for causes of the same situation (e.g., "given that their memory is less effective, how likely is it they drink alcohol?"). Importantly, the two types of reasoning conceptually follow a cause-effect order in one case (i.e., predictive; alcohol-memory loss) and an effect-cause order in the other case (i.e., diagnostic; memory loss-alcohol). Notably it has been well supported that the two types of reasoning are not symmetric: predictive (vs. diagnostic) reasoning is the preferred way and leads to stronger perceived causal relations because it matches the natural order of events, with causes preceding effects (e.g., Tversky & Kahneman, 1974; Barr, 2010; Fenker, Waldmann, & Holyoak, 2005).

How is this linked to word order? Indeed, only one research to our knowledge considered the effect of both reasoning types (predictive and diagnostic) next to the order in which causes and effects were presented. Bettinsoli and colleagues (2020) argued that the perceived relation between causal elements (e.g., alcohol and memory) is not only limited to the type of reasoning people may engage in, but that it can also be influenced by the order in which causes and effects appear in a sentence (i.e., alcohol-memory or memory-alcohol). Across different methodologies (e.g., correlational or experimental), methodologies (e.g., role playing or solving riddles), and measures (e.g., coding and decoding information), the authors consistently showed that both order and type of reasoning contribute to the perceived causal relations between elements. Specifically, the cause-effect order in predictive reasoning (e.g., "given that they often drink alcohol, it is likely that their memory is less effective") was stronger than the other order and type of reasoning combinations (e.g., diagnostic reasoning following a cause-effect order; "it is likely that they drink alcohol, given that their memory is less effective") when it comes to perceive the causal relation between two events, elements, or sentences.

Importantly, there are reasons to believe that there is much applied value in order-effect research, especially with a focus on a specific domain where causal reasoning is critical—the health domain. For instance, in line with binomial research presented before, Offringa and colleagues (2019) focused on the specific word order in a common binomial message, "Fruit and Vegetables", and found that the American daily intake of vegetables (commonly mentioned second) was lower compared to fruits. Therefore, the authors suggested that future messages should purposely

put "vegetables" first in order to emphasize their importance regarding contribution to health and ultimately enhance their intake.

Another direct application comes from a more recent set of studies by Bettinsoli and Suitner (2022), which focused on health-related issue and the way in which conveying messages might impact on intention to change unhealthy behaviors. Specifically, by considering unhealthy habits as causes (i.e., smoking) and health-related outcomes as effects (i.e., lung cancer), they investigated whether there is a specific word ordering that campaign planners should use in their attempts to persuade people to change their behaviors. Typically (but not exclusively, Bettinsoli et al., 2020) effect-cause order matches the diagnostic reasoning (e.g., "given that you have stomachache, it is likely that you ate something toxic") and the compliance to a medical prescription requires a diagnostic reasoning (e.g., "given that you have some health problems, you should quit what is causing it"). Based on these assumptions, Bettinsoli and Suitner (2022) predicted -and found- that an effect-cause order (e.g., "given that you have high glycemia [effect], you should reduce sugars intake [cause]") is more suitable than a cause-effect order (e.g., "you should reduce sugars intake, given that you have high glycemia") in messages encouraging health behaviors. It might be that the effect-cause order may promote a focus on the negative effects, which may be particularly potent to trigger the motivation to manage and reduce the sources of a symptom by adopting the prescribed action. Thus, in the medical communication realm -and in promotion and prevention campaigns in general- an effect-first order may be a more effective way to communicate health recommendations to promote a higher compliance with prescribed behaviors.

Conclusions

To conclude, this is not to claim that order is the only – or even the main- critical variable to consider while studying both production and comprehension of language, and more work is needed to identify which other cognitive and attentional processes might be involved in processing causality and information more generally. Of course, this was not meant to be an exhaustive review on all the research concerning the effects of word order because much of this research has been related to linguistics domain. However, this chapter illustrated how such a subtle feature of language might be relevant while studying attributional and inferential processes, causal relations, and intentional behaviors. Importantly, first

basic research was needed to better understand the underlying mechanisms of word order effects on a higher cognitive level. Now, time is ripe to include additional research, which might be essential to overcome some limits that previous research have encountered, but especially to extend word order studies to applications in different domains and processes, such as for example stereotyping and stereotype reductions, where the order in which information about an individual appear might have a crucial role in either trigger or reduce stereotypical ascriptions.

Final notes

As a final – and personal- note, I would like to thank Anne Maass for being such a stratospheric PhD mentor, and no doubt one of the most stimulating persons I've ever met. In fact, if I'm here today writing about word order, and more generally, doing research, it is because it's rare to find someone who can inspire you so deeply as Anne did with me (and with many authors of this book, I guess). Keeping it on ordering and causal relations: Anne sparked my passion for research. Ironically, I'm writing this chapter from the office –and the exact same desk– that used to be hers before she moved to NYU Abu Dhabi. It's the desk where everything started. I moved out for a while, and then I moved back. I can't hide that sitting at -what used to be- Anne's desk is a big responsibility and I'm afraid I would ever be as good as her at this job. However, I can say that, while crossing the world up and down and from left-to-right and back, I made –or I eventually found– my own order.

References

Asch, S. E. (1946). Forming impressions of personality.The Journal of Abnormal and Social Psychology, 41, 1230–1240.

Bader, M., & Bayer, J. (2006). Case and linking in language comprehension: Evidence from German (Vol. 34). Springer Science & Business Media.

Bader, M., & Meng, M. (1999). Subject-object ambiguities in German embedded clauses: An across-the-board comparison. Journal of Psycholinguistic Research, 28(2), 121-143.

Bates, E., & MacWhinney, B. Functionalist approaches to grammar. In L. Gleitman & E. Wanner (Eds.), Language acquisition: The state of the art. New York: Cambridge University Press, 1982.

Bentz, C., & Christiansen, M. H. (2010). Linguistic adaptation at work?

The change of word order and case system from Latin to the Romance languages. In Proceedings of the 8th International Conference on the Evolution of Language (pp. 26-33).

Bettinsoli, M. L., & Suitner, C. (2022). BeCause of the Effect the role of health messages ordering on behavioral change intention. Journal of Psycholinguistic Research, 51(3), 563-576.

Bettinsoli, M. L., Maass, A., & Suitner, C. (2019). The first, the least and the last: Spatial asymmetries in memory and their relation to script trajectory. Memory & Cognition, 47, 229-239.

Bettinsoli, M. L., Maass, A., Kashima, Y., & Suitner, C. (2015). Word-order and causal inference: The temporal attribution bias. Journal of Experimental Social Psychology, 60, 144-149.

Bettinsoli, M. L., Suitner, C., & Maass, A. (2020). The distinct contributions of cause–effect order and reasoning type in judgments of causality. Journal of Cognitive Psychology, 32(1), 108-129.

Bettinsoli, M. L., Suitner, C., Maass, A., Finco, L., Sherman, S. J., & Salvador Casara, B. G. (2022). The spatial ingroup bias: Ingroup teams are positioned where writing starts. Personality and Social Psychology Bulletin, 48(1), 49-64.

Bohner, G. (2001). Writing about rape: Use of the passive voice and other distancing text features as an expression of perceived responsibility of the victim. British Journal of Social Psychology, 40(4), 515-529.

Bornkessel, I., Zysset, S., Friederici, A. D., Von Cramon, D. Y., & Schlesewsky, M. (2005). Who did what to whom? The neural basis of argument hierarchies during language comprehension. Neuroimage, 26(1), 221-233.

Boroditsky, L. (2001). Does language shape thought?: Mandarin and English speakers' conceptions of time. Cognitive psychology, 43(1), 1-22.

Bruckmüller, S., Hegarty, P., & Abele, A. E. (2012). Framing gender differences: Linguistic normativity affects perceptions of power and gender stereotypes. European Journal of Social Psychology, 42(2), 210-218.

Chomsky, N. (1986). Knowledge of language: Its nature, origin, and use. Greenwood Publishing Group.

Cooper, W. E., & Ross, J. R. (1975). World order. Papers from the parasession on functionalism, 11, 63-111.

Dröge, A., Fleischer, J., Schlesewsky, M., & Bornkessel-Schlesewsky, I. (2016). Neural mechanisms of sentence comprehension based on predictive processes and decision certainty: electrophysiological

evidence from non-canonical linearizations in a flexible word order language. Brain research, 1633, 149-166.

Dryer, M. S. (2011). Order of Subject, Object, and Verb. The World Atlas of Language Structures, 330-333.

Fenker, D. B., Waldmann, M. R., & Holyoak, K. J. (2005). Accessing causal relations in semantic memory. Memory & cognition, 33, 1036-1046.

Gattei, C. A., París, L. A., & Shalom, D. E. (2021). Information structure and word order canonicity in the comprehension of Spanish texts: an eye-tracking study. Frontiers in Psychology, 12, 629724.

Gernsbacher, M. A., Goldsmith, H. H., & Robertson, R. R. (1992). Do readers mentally represent characters' emotional states?. Cognition & Emotion, 6, 89-111.

Gibson, E., Piantadosi, S. T., Brink, K., Bergen, L., Lim, E., & Saxe, R. (2013). A Noisy-Channel Account of Crosslinguistic Word-Order Variation. Psychological Science, 24, 1079-1088.

Givón, T. (1992). On interpreting text-distributional correlations: Some methodological issues. In D. L. Payne (Ed.), Pragmatics of word order flexibility (pp. 305-320). Philadelphia: Benjamins.

Goldin-Meadow, S., & Alibali, M. W. (2013). Gesture's role in speaking, learning, and creating language. Annual Review of Psychology, 64, 257-283.

Goldin-Meadow, S., So, W. C., Özyürek, A., & Mylander, C. (2008). The natural order of events: How speakers of different languages represent events nonverbally. Proceedings of the National Academy of Sciences, 105, 9163-9168.

Gómez, R. L., & Gerken, L. (2000). Infant artificial language learning and language acquisition. Trends in cognitive sciences, 4(5), 178-186.

Greenberg, J. H. (1963). Some universals of grammar with particular reference to the order of meaningful elements. Universals of Language, 2, 73-113.

Grüning, A. (2003). Why Verb-Initial Languages are not Frequent. MPI-MIS Preprint Series,10. Max Planck Institute for Mathematics in the Sciences, Leipzig.

Hall, M. L., Mayberry, R. I., & Ferreira, V. S. (2013). Cognitive constraints on constituent order: Evidence from elicited pantomime. Cognition, 129, 1-17.

Hamilton, D. L., & Sherman, S. J. (1996). Perceiving persons and groups. Psychological Review, 103, 336-355.

Haspelmath, M. (2021). Explaining grammatical coding asymmetries: Form–frequency correspondences and predictability. Journal of

Linguistics, 57(3), 605-633.

Hegarty, P. (2015). Ladies and gentleman: Word order and gender in English. In G. G. Corbett (Ed.), The expressions of gender (Vol. 6, pp. 69-86). Berlin, Germany: de Gruyter Mouton.

Hegarty, P., Lemieux, A. F., & McQueen, G. (2010). Graphing the order of the sexes: Constructing, recalling, interpreting, and putting the self in gender difference graphs. Journal of Personality and Social Psychology, 98, 375-391.

Hegarty, P., Mollin, S., & Foels, R. (2016). Binomial word order and social status. In H. Giles & A. Maass (Eds.), Advances in intergroup communication (pp. 119-135). New York, NY: Peter Lang.

Hegarty, P., Watson, N., Fletcher, L., & McQueen, G. (2011). When gentlemen are first and ladies are last: Effects of gender stereotypes on the order of romantic partners' names. British Journal of Social Psychology, 50, 21-35.

Johnson-Laird, P. N. (1983). Mental models: Towards a cognitive science of language, inference, and consciousness. Cambridge, MA: Harvard University Press.

Johnson, S. A., & Braber, N. (1998). Exploring the German language. London: Arnold Publishers.

Kaiser, E., & Trueswell, J. C. (2004). The role of discourse context in the processing of a flexible word-order language. Cognition, 94(2), 113-147.

Lamers, M., & De Swart, P. (Eds.). (2011). Case, word order and prominence: Interacting cues in language production and comprehension (Vol. 40). Springer Science & Business Media.

Langus, A., & Nespor, M. (2010). Cognitive systems struggling for word order. Cognitive Psychology, 60, 291-318.

Lupyan, G., & Christiansen, M. H. (2002). Case, word order, and language learnability: insights from connectionist modeling. In Proceedings of the 24th Annual Conference of the Cognitive Science Society (pp. 596-601).

Maass, A., Pagani, D., & Berta, E. (2007). How beautiful is the goal and how violent is the fistfight? Spatial bias in the interpretation of human behavior. Social Cognition, 25(6), 833-852.

Maass, A., Suitner, C., & Deconchy, J. P. (2014). Living in an Asymmetrical World: How Writing Direction Affects Thought and Action. Psychology Press.

Maass, A., Suitner, C., & Nadhmi, F. (2014). What drives the spatial agency bias? An Italian–Malagasy– Arabic comparison study. Journal of

Experimental Psychology: General, 143(3), 991.

MacWhinney, B., 1977. Starting points. Language, 53, 152-168.

McDonald, J. L., Bock, K., & Kelly, M. H. (1993). Word and world order: Semantic, phonological, and metrical determinants of serial position. Cognitive Psychology, 25(2), 188-230.

McGuire, W. J., & McGuire, C. V. (1982). Significant others in self-space: Sex differences and developmental trends in the social self. Psychological Perspectives on the Self, 1, 71-96.

Mollin, S. (2012). Revisiting binomial order in English: ordering constraints and reversibility. English Language and Linguistics, 16(01), 81-103.

Offringa, L. C., Stanton, M. V., Hauser, M. E., & Gardner, C. D. (2019). Fruits and vegetables versus vegetables and fruits: rhyme and reason for word order in health messages. American journal of lifestyle medicine, 13(3), 224-234.

Pratto, F., Hegarty, P. J., & Korchmaros, J. D. (2007). How communication practices and category norms lead people to stereotype particular people and groups. Stereotype dynamics: Language based approaches to the formation, maintenance, and transformation of stereotypes, 293-313.

Rozin, P., Berman, L., & Royzman, E. (2010). Biases in use of positive and negative words across twenty natural languages. Cognition and Emotion, 24(3), 536-548.

Sandler, W., Meir, I., Padden, C., & Aronoff, M. (2005). The emergence of grammar: Systematic structure in a new language. Proceedings of the National Academy of Sciences of the United States of America, 102(7), 2661-2665.

Segal, E.M. (1995). Cognitive-phenomenological theory of fictional narrative. In J.F. Duchan G.A. Bruder, & L.E. Hewitt (Eds.), Deixis in narrative: A Cognitive Science Perspective, (pp. 61-78). Hillsdale, NJ: Erlbaum.

Song, J. J. (2001). Linguistic Typology: Morphology and Syntax. Pearson Education.

Suitner, C., Maass, A., Navarrete, E., Formanowicz, M., Bratanova, B., Cervone, C., ... & Carrier, A. (2021). Spatial agency bias and word order flexibility: A comparison of 14 European languages. Applied Psycholinguistics, 42(3), 657-671.

Tabullo, Á., Arismendi, M., Wainselboim, A., Primero, G., Vernis, S., Segura, E. Zanuto, & Yorio, A. (2012). On the learnability of frequent and infrequent word orders: An artificial language learning study. The Quarterly Journal of Experimental Psychology, 65, 1848-1863.

Tily, H., Frank, M. C., & Jaeger, T. F. (2011). The learnability of constructed languages reflects typological patterns. In Proceedings of the 33rd Annual Conference of the Cognitive Science Society (pp. 1364-1369).

Tomlin, Russell S. Basic Word Order (RLE Linguistics B: Grammar): Functional Principles. Vol. 13. Routledge, 2014.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. Science, 185(4157), 1124-1131.

Zwaan, R. A., & Radvansky, G. A. (1998). Situation models in language comprehension and memory. Psychological Bulletin, 123, 162-185.

9. Spatial Agency Bias: mapping social agency into the visual field

Caterina Suitner[1] *and Magdalena Formanowicz*[2]
[1]University of Padova, Italy
[2]SWPS-Warsaw, Poland

Writing and reading are embedded in our everyday life. We read books, instructions, messages, signs, and the news; we write notes, emails, and status updates on Twitter or Facebook. This list could be very long and comprise many of our daily routines, yet we often do not realize how frequently we are involved in such activities, and that they can have unexpected effects on how we think about social world. Writing and reading entail eye and, for writing, also hand movements that are systematically directed. For example, in the present text, the direction is from left toward right, because we are writing in English. If we were writing in Arabic, the direction would have been the opposite. In this chapter, we are going to inspect the role of writing and reading habits for social cognition, to show that direction of texts (e.g., rightward for English and leftward for Arabic) creates a spatial model for thinking. The direction of script establishes a clear spatial organization of information, including socially relevant information, allowing for the creation of shared reality. This can be compared to the railways, on which the mental images of our social world travel in our mind and in the mind of others, as people in the same culture share the journey direction. One key aspect of this process is that this spatial organization of mental images is not devoid of meaning: The trajectory imposed by textual data affects how we mentally envisage action in general but also is enriched by social content, and specifically it is associated with agency, a key dimension of social thinking.

Briefly, we can define agency as being related to goal-oriented action, dynamism and striving.

Although concise, such a definition already suggests how fundamental agency is for individuals and groups. The need of accomplishing goals is not only crucial for one's survival (New et al., 2007; Frith & Frith, 2010), but also for the self and group fulfillment (Bandura, 2001). Given the importance of agency for social cognition, we prioritize cues of agency. Accordingly, the Spatial Agency Bias (SAB, Suitner & Maass, 2016) posits that we spatially encode social agency along the reading/writing direction. Furthermore, the direction through which we encode social targets is a subtle cue of social roles we assign to members of given groups in relation to stereotype content. For example, in Western cultures, we reserve the agentic left position facing rightward to members of high-status groups (e.g., men) in visual representations. Adam and Eve are generally portrayed with Adam to the left and Eve to the right (Maass, Suitner, Favaretto, & Cignacchi, 2009). Spatial images, therefore, reveal how we think about social targets and represent a valuable instrument to inspect social cognition in a subtle and not invasive way. The chapter will first explore some specific linguistic features that together with text direction delineate the general spatial asymmetry in information processing, then address the cognitive process behind the switch from a general schema for action to Spatial Agency Bias, and finally focus on its concrete implications for social relations.

Linguistic Underpinnings of SAB

Scanning and writing activities create a generalized mental schema for action. Immediate evidence of this generalization can be easily achieved by observing how the spatial direction of the language emerges in contexts that have nothing to do with reading or writing (Maass, Suitner, & Dechonchy, 2014). For example, when people organize objects on a shelf, they are likely to create a spatial pattern that sets the biggest object to left and the smallest to right if they are socialized in a culture whose language is written from left to right. In contrast, they are more likely to create right-to-left patterns if they usually read and write in Hebrew, Arabic, or Urdu. Indeed, script direction has been proven to be relevant in many contexts, including how we mentally represent numbers (Zebian, 2005), time (Boroditsky, 2001), in our drawings (Singh and Vaid, 1987), and in art production (Pérez González, 2012). It also affects our aesthet-

ic preferences, as we favor representations that match the direction of our language (Chokron & De Agostini, 2000; Chahboun, Flumini, Pérez González, McManus, & Santiago, 2017). For example, football fans appreciate a soccer goal more if it is spatially evolving along their habitual script direction rather than if it is going the 'wrong' way (Maass, Pagani, & Berta, 2007).

However, the contribution of language to SAB is not limited to this script directionality. Language is one of the richest nourishments to feed and build our social reality and offers a plethora of ingredients that are mixed in secret recipes, resulting in a rich taste palette. The meaning of the words is just one of the ingredients, together with the grammatical categories, the word order, the syntax, the pragmatics, and the nonverbal context. Accordingly, mechanisms of SAB were related to two language-related features beyond writing direction, namely the ordering in basic transitive sentences of agent and object, and the flexibility of word ordering.

Starting with the first feature, namely the ordering of the agent, it is very easy to disentangle the agent (A) and object (O) thematic roles in a standard sentence such 1. We can all quickly agree that Anne is the agent and Caterina the object of the transitive sentence and, given that we are in the English rightward context, Anne is to left of Caterina (A → O). Thanks to Google Translate, we have also managed to have the same sentence in Hebrew (2.) and in Malagasy (3.).

1. Anne helps Caterina

   A → O

2. אן עוזר לקטרינה

   O ←A

3. Manampy an'i Caterina i Anne

   O → A

In all the three versions Anne has been marked in bold (Anne, אן, and Anne, respectively) and Caterina in italic (Caterina, לקטרינה, and Caterina, respectively). In the second version, Hebrew, the text is organized from right to left, and we can see that Anne ends up being to the right (O ←A). In line with the spatial agency bias model, among Hebrew speakers, the agent is associated to right and the action is evolving toward left. So far, this is congruent with the writing direction effect we have mentioned. Things get complicated when we consider the third example. Malagasy is read from left to right (like English), so the direction is rightward. However, the Malagasy word order places the agent after the object (O → A).

Because of this word order, Anne appears to right in the written sentence, like in the Hebrew version. This very peculiar (and fascinating) structure, present in only 3% of languages (Dryer, 2013), allows for investigating the joint role of script direction and word order in SAB. We investigated this by comparing Arabic (O ←A), Italian (A → O), and Malagasy (O → A) respondents in two tasks (Maass, Suitner & Nadhmi, 2014). One was a drawing task, and since only the focal object was mentioned ("draw a fish"), the order of words was not very salient. In this task, all participants arranged the focal object in their drawing according to their writing direction, namely, moving rightward for Italian and Malagasy and leftward for Arabic respondents. In the other task, we asked participants to choose which of the pair mirror versions of a drawing representing two targets better represented the given interaction (similar to Anne helps Caterina). The mirror drawings arranged the agent and the object on the horizontal vector so that in one vignette the agent was to the left of the object (AO), and in the other to the right (OA). In this task, the preferred ordering of the agent and the object in the syntax defined the spatial arrangement of agent and object in the language, with Italian respondents preferring AO and both Arabic and Malagasy respondents preferring OA. This study shows that both script direction and word ordering are relevant features for the emergence of a spatial asymmetry, and that they can be contextually activated by the task.

Agent and object ordering sometimes can also vary within a language, such as in passive sentences (e.g., Caterina is helped by Anne), where the object is mentioned before the agent because of its relevance (Payne, 1992). The active form is considered the more basic form (typical, normal, frequent, easier, unmarked), whereas the use of passive voices marks specific communicative needs, including signaling the importance of the logical object (Anisfeld & Klenbort, 1973). This pragmatic relevance stressed by OA ordering is indeed captured by participants, who reduce or even invert the SAB in passive sentences (Halicki, Suitner, Vogel, & Wänke, 2021), confirming the importance of word ordering for assigning relevance to a target (Kesebir, 2017; see also Bettinsoli's Chapter 8 in the present book). This leads to the last linguistic feature that has been so far linked to the SAB, namely linguistic flexibility. Indeed, some languages, such as English, have a very rigid word order, whereas in others it is easier to shuffle the words of the sentence without threatening the intelligibility of its meaning (e.g., Siewierska, Rijkhoff, & Bakker, 2010). Presence of higher flexibility is typically related to linguistic features that mark

thematic roles (e.g., through suffixes marking syntactic roles or through overt person–number agreement) and hence reduce the ambiguity in the assignment. We investigated the role of language flexibility asking speakers of 14 different languages (all written from left to right) to draw an interaction between two persons (Suitner et al. 2021). We then coded both the flexibility of the language and whether the interaction was drawn to be in line with the SAB (with agent to the left). Participants whose language is characterized by a rigid syntax were more likely to display the SAB compared to participants whose language had higher word order flexibility.

To summarize, each of the presented factors, namely the role of word ordering in the syntax, the role of contextual activation of AO or OA orders due to specific pragmatic choices (e.g., using the passive form), and the flexibility of ordering within a language, interact with writing direction (a generally stable linguistic feature), often strengthening SAB, but sometimes also limiting its presence. Further studies are needed to disentangle the role of agent-recipient positioning and of action direction in the SAB, and their relationship with contextual features.

From text direction to social agency

Agency is considered by many scholars a basic dimension of social stereotypes (Spence et al., 1978; see also Formanowicz & Suitner's chapter on 7 in this book). Even if agency seems like a very intuitive concept referring to "goal achievement and task functioning" (Abele & Wojciszke, 2014, p. 197), a shared and specific definition is still lacking in the scientific community, and this is reflected for example in a variety of measurement issues (Cavazzoni et al., 2022. Until a unique and shared wording for defining agency is reached, the different aspects stressed by different scholars provide important theoretical inspirations for better understanding the processes behind the Spatial Agency Bias. A to-the-point definition has been provided by an anthropologist, Ahearn (2001), according to which "Agency refers to socio-culturally mediated capacity to act" (p. 112). This definition stresses the role of the social context to construe the concept of agency and is particularly suitable to address the Spatial Agency Bias, which transforms a habitual *act* (e.g., writing) into higher-order *social* cognition.

The process behind the transformation of script direction into the trajectory of mental representation of agency can be theoretically framed

within an embodied perspective (Suitner & Maass, 2016; Hegarty, in press), which stresses the situated and action-oriented nature of cognition and the role of bodily experiences and sensory-motor stimulations for higher-order functioning (Barsalou, 2008). The basic idea of embodiment is that the body matters and is an integral component of cognition. According to a strong embodied perspective, cognitive processes are grounded in the body even when perceptions and bodily states are not *hic et nunc* stimulated by the external environment. Therefore, bodily states remain relevant even offline, as they are interiorized and simulated so that they become integral part of cognition (Wilson, 2002). The oriented eye movements performed during the visual scanning of texts and the trajectories executed with our arm during handwriting activity offer culture-specific motoric stimulations. Accordingly, the idea of a culturally embodied cognition (Bettinsoli, Suitner & Maass, 2021) suggests that the culturally defined convention of the writing system exerts a systematic effect on how we envisage and perform action, including mental actions, which are therefore situated in the mental and actual space.

This mental schema is present at different stages of cognition. At the basic level of categorization, a spatial asymmetry can be observed in the very fundamental task of distinguishing targets belonging vs. not belonging to a given category. For example, German respondents were faster in categorizing members of a given category when they were positioned to the left than to the right in a visual field (von Hecker & Klauer, 2021).

The schema also intrudes in our attention, so that we process information differently according to how it is displayed in a visual field: there is a cognitive advantage to information that is visually positioned at the starting point of the writing trajectory (in terms of attention and memory, Bettinsoli, Maass, & Suitner, 2019; Mendonça, Garrido, & Semin, 2020a) and to information that is arranged consistently with script trajectory (in terms of decision making and time for processing, Mendonça, Garrido, & Semin, 2022, in terms of attention fluency, Spalek & Hammas, 2005). The primacy effect of the spatial position related to the habitual script affects also inferences, so that targets presented at the starting position (i.e., to the left for English participants) are attributed more of any given quality (von Hecker et al., 2022), including of course agency (e.g., Mendonça, Garrido, & Semin, 2020b).

At the encoding level, how we represent the agent and the recipient of an action in a visual field (see also seminal observations on thematic role assignment in aphasic patients in Chatterjee, Maher, & Heilman,

1995), is also guided by script direction, with right-to-left readers envisaging the agent to right and the recipient to the left, left-to-right readers vice versa (Maass & Russo, 2003). The specific role of directed movements in such representations has been tested by involving participants in a writing exercise aimed at training them in different writing directions (Suitner, Maass, Bettinsoli, Carraro, & Kumar, 2017). The repetition of directed movements did indeed affect how people then spatially represented social interactions, confirming the motor component of the SAB. Together, these findings suggest that situating action in a horizontally evolving mental space results in a general and shared mental schema for agency that matches script direction. These results also stress the role of culture in shaping these habits, as spatial bias is influenced by repeated behavioral experiences, namely the momentary experience with the 'opposite' writing direction.

The social implications and applications of Spatial Agency Bias

Earlier on we introduced agency as a fundamental dimension of social cognition, and we now review the relevance of the SAB for social cognition at both the encoding and decoding level. On the one hand, the SAB reveals socio-cognitive processes, as targets that are perceived as agentic are more likely to be represented in line with script directionality. On the other side, the direction used to represent social targets conveys agency, therefore communicating and reproducing stereotype content.

With respect to the encoding level, the stereotype content is revealed through space. For example, the content of religious event, which are surely not anchored to any specific experience, are highly affected by spatial schema that contribute to the meaning making of such mysterious and unknown contents. Mc Manus' review of Italian renaissance art (2005) offers several instances, such as the systematic rightward movement of the angel in Annunciation's representations. Also members of groups that are stereotyped as high in agency, including men (vs. women) and young (vs. elderly) people, are more likely to be envisaged along the script trajectory. For example, a Gender Spatial Bias has been reported in several domains, as men and women are differently represented in the horizontal vector both in experimental tasks (Maass, et al. 2009), in artwork representations (Humphrey & McManus, 1973; Gordon, 1974; Grüsser et al., 1988; ten Cate, 2002) or in graphs portraying gender differences in scientific journals (Hegarty & Buechel, 2006). Also, this coupling

of men with the agentic position is specifically driven by the higher status they are assigned (Carnaghi, Piccoli, Brambilla, & Bianchi, 2014). In line with this interpretation, and with the concept that agency is attributed to the self (Abele & Wojciszke, 2007), the analysis of the direction of portraits by Rembrandt led Mc Manus (2005) to conclude that the right cheek is associated to the self, being prevalent in portraits of self and male kins more than in portraits of female non-kin sitters. Along the same line are the asymmetries observed in portraits of women and men sitter (Suitner & Maass, 2007) which reported decline over time of the gender bias that reserved the rightward direction to male sitters. This suggests a match between a spatial schema for gender and a social schema for roles, pointing at a role of spatial visualization for common social meaning. If shared content is expressed through space, can we also affirm the opposite path, namely that spatial layouts hint the congruent world view? Even if the evidence of this second path is less robust, some studies provided preliminary support for a role of spatial cues in information processing.

At the decoding level, the direction of script is associated to agency, so that rightward oriented (from the viewer's perspective) face profiles are perceived as more agentic by observers whose language is also written rightward (Mendonça, Garrido, & Semin, 2020b; Suitner et al. 2017). Despite the general conflation with positive aspects, the attribution of agency can also have negative consequences specifically when the agency implies responsibility for a negative event (Fausey & Boroditsky, 2011). Along this line of reasoning, Halicki, Hauser, and Wänke (2022) showed that German participants blamed female targets more for a sexual harassment when the targets were visualized to the left than to the right side of a picture. The effect was associated with the levels of acceptance of sexism and rape myths of the participants, showing that bias is a subtle symptom to diagnose the beliefs of the participants.

Importantly, social information is more fluently elaborated when it is spatially congruent with the content of the stereotype. For example, rightward (vs. leftward) profiles of male (vs. female) target are categorized faster by Italian participants (Suitner et al. 2017, and advertisements of products whose direction matches the stereotype content (i.e., masculine products moving from left to right in a Western culture) are more likely to enhance the trust in the advertised brand (Monahan & Romero, 2020). On the other hand, when the spatial encoding does not match the stereotype content, it can represent a system threat, challenging the belief in the "way things are", and participants can be willing to reaffirm

the social order. This was shown by Italian respondents in a study by Mazzurega et al. (2019): participants holding a sexist world view were less likely to attribute agentic jobs to a woman presented facing rightward (vs. leftward) to the extent they perceived the rightward face profile as gender-atypical for the target. This not only matches the reported finding a temporal attenuation of gender spatial bias in art (Suitner and Maas, 2007), but also the finding by Hegarty (in press) on scientific journals, where the androcentric convention to present men's data to the left and women's data to right in tables and graphs is slowly weakening and is related to feminist ideology.

Applied to the intergroup realm, the SAB appears to be relevant in social categorization and group discrimination, as the agentic script trajectory advantages the ingroup. This is in line with the idea that agency is a desired characteristic (Abele & Wojciszke, 2014; Suitner & Maass, 2008) and is especially relevant in contexts where being goal oriented is literally at stake, such as a soccer match, where you want your team to win by scoring goals. This optimal context of investigating SAB was applied to study the idea that favoritism for the own team (or ingroup bias, using the vocabulary of the Social Identity Theory, Tajfel & Turner, 1982) will result in a spatial bias. Indeed, Italian participants set the soccer team in the visual field so that their own team was playing in a left-to-right trajectory, scoring to the right. The Arabic-speaking participants created the opposite spatial layout.

Altogether, the reported findings suggest that the spatial schema for agency has many social implications, as it can be exploited to diagnose social stereotypes, favoritism of the own group, and discrimination of targets that do not match social expectations.

Conclusions and future directions

The SAB is a phenomenon that the first author was lucky enough to investigate during her Ph.D. under the supervision of Anne Maass. It is a tiny field of study, and it is now a pleasant surprise to read works from other laboratories and to see that our understanding of the phenomenon is increasing. Sometimes the effect sizes of the SAB are small, however, its pervasiveness and coherence, together with its subtle but ubiquitous consequences make it a relevant phenomenon. Moreover, given that our experience is constantly embedded in space, a rigid and strong SAB would not allow us to interact flexibly with our physical and social

environment. As a consequence, the subtlety and context-dependance of this horizontal spatial bias is not only unsurprising, but also functional.

Excitingly, the range of possible consequences and applications in the field of social cognition and mass communication is wide, and further studies are needed to fully explore the potential of the SAB. For example, little is known about the feasibility of using space in interventions aimed at promoting social equality. Tentative evidence comes from an experiment in which we trained participants to associate male and female targets in the direction opposite to the gender stereotype. Overexposing Italian participants to rightward female profiles and leftward female profiles (compared to the opposite pattern) reduced their subsequent levels of sexism (Suitner et al., 2017). More empirical efforts are needed to replicate and expand this preliminary effort.

The role of SAB for self-presentation could also be further explored, especially in the current moment, when selfies are very common. Does the profile chosen in a selfie match the self-presentation motives in terms of agency? First evidence comes from a study by Nicholls, Clode, Wood, and Wood (1999) in which participants, regardless of gender, presented their left cheek when asked to pose for a family portrait and their right cheek when posing as scientists. More studies are needed to explore the spatial asymmetries in big data, such as social networks. Are gender differences in the direction of face profiles moderated by the gender inequality at the country level?

In terms of our theoretical understanding of the SAB, its embodied nature deserves further inspection, as so far we have only indirect evidence of the involvement of our motoric system in the SAB. Along the same line, further interdisciplinary studies are needed to understand the interplay between the contextual factors and neurocognitive underpinning of the bias, which were investigated from independent approaches. Indeed, we cannot exclude interactive effects between the culture and the neurological architecture and dynamic or distal neurological effects. In fact, we have still not fully understood why the left-to-right direction is so overwhelmingly prevalent among the writing systems.

Finally, on a more basic level, SAB and other phenomena related to the privileged processing of subtle cues related to agency (see also chapter of Formanowicz & Suitner, this book; Weis et al., 2022) can be relevant to the understanding of the process of cultural transmission. It is quite fascinating, that a very basic and evolutionary founded attunement to cues of activity and agency (framed also as intentionality) is embedded

in language and in processing of spatial information. On the one hand, it is reflecting the adaptive role of attending to cues of agency, because other agents can affect our own actions. Furthermore, to pursue various goals, we need to interact with others and form meaningful coalitions that allow for joint actions (Sebanz et al., 2006). The ability to encode and decode action(agency)-oriented stimuli easily, can help us in communication leading to such a coordinated action. On the other hand, in a social world, the position of an agent often coincides with social hierarchies (see Peter Hegarty's Chapter 10 in this book), and language and spatial arrangements not only reflect this association, but also strengthen it through the process of cultural transmission (see Yoshisha's Kashima's Chapter 1 in this book). By bringing these phenomena to a wider audience, we can possibly correct for their automatic effects.

We would like to dedicate this chapter to Anne Maass, who of course is central for the SAB literature, being the pioneering author of the first paper on this topic (Maass & Russo, 2003), and co-author of many of the subsequent work.

Also, Anne is the person that connected the two authors, and inspires a collaboration style that is spreading in a snowball to the many researchers around the world that were lucky enough to meet her, many also involved in this book. Her approach prompts us toward a broader picture, in which the whole is greater than the sum of single isolated effects and individual researchers. For the benefit of both science and scientists.

References

Abele, AE, & Wojciszke, B. (2007). Agency and communion from the perspective of self versus others. Journal of Personality and Social Psychology, 93, 751–763. Social Psychology, 93, 751-763.

Ahearn, L. M. (2001). Language and agency. Annual review of anthropology, 109-137.

Anisfeld, M. & Klenbort, I. (1973). On the functions of structural paraphrase: The view from the passive voice. Psychological Bulletin, 79(2), 117.

Bandura, A. (2001). Social cognitive theory: an agentic perspective. Annual Review of Psychology, 52, 1–26. https://doi.org/10.1146/annurev.psych.52.1.1

Barsalou, L. W. (2008). Grounded cognition. Annual Review of Psychology, 59, 617-645. https://doi.org/10.1177/10778012221108420

Bettinsoli, M. L., Maass, A., & Suitner, C. (2019). The first, the least and the last: Spatial asymmetries in memory and their relation to script trajectory. Memory & Cognition, 47(2), 229-239.

Bettinsoli, M. L., Suitner, C., & Maass, A. (2021). Take a Walk on the Cultural Side: A Journey into Embodied Social Cognition. Handbook of Embodied Psychology, 423-450.

Boroditsky, L. (2001). Does language shape thought?: Mandarin and English speakers' conceptions of time." Cognitive psychology 43(1), 1-22.

Carnaghi, A., Piccoli, V., Brambilla, M., & Bianchi, M. (2014). Gender hierarchy in the space: The role of gender status in shaping the spatial agency bias. The Journal of social psychology, 154(2), 105-114.

Cavazzoni, F., Fiorini, A., & Veronese, G. (2021). How do we assess how agentic we are? A literature review of existing instruments to evaluate and measure individuals' agency. Social Indicators Research, 1-29.

Chahboun, S., Flumini, A., Pérez González, C., McManus, I. C., & Santiago, J. (2017). Reading and writing direction effects on the aesthetic appreciation of photographs. Laterality: Asymmetries of Body, Brain and Cognition, 22(3), 313-339.

Chatterjee, A., Maher, L. M., & Heilman, K. M. (1995). Spatial characteristics of Asymmetries Drive Spatial Attention. Cognitive Science, 46(8).

Chokron, S., & De Agostini, M. (2000). Reading habits influence aesthetic preference. Cognitive Brain Research, 10(1-2), 45-49.

Dryer, M. S. (2013). Order of Subject, object and verb. In Dryer, M. S., & Haspelmath, M. (Eds.), The world atlas of language structures online. Leipzig: Max Planck Institute for Evolutionary Anthropology.

Fausey, C. M., & Boroditsky, L. (2011). Who dunnit? Cross-linguistic differences in eye-witness memory. Psychonomic Bulletin & Review, 18(1), 150–157.

Frith, U., & Frith, C. (2010). The social brain: allowing humans to boldly go where no other species has been. Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences, 365(1537), 165–176. https://doi.org/10.1098/rstb.2009.0160.

Gordon, I. (1974). Left and right in Goya's portraits. Nature, 294, 197-198.

Grüsser, O. J., Selke, T., & Zynca, B. (1988). Cerebral lateralization and

some implications for art, aesthetic perception and artistic creativity. In I. Rentschler, B. Herzberg, & D. Epstein (Eds.), Beauty and the brain: Biological aspects of aesthetics (pp. 257-293). Boston: Birkhauser

Halicki, K. T., Hauser, R., & Wänke, M. (2022). When she is Standing Left, she Might be Blamed. Responsibility Attribution for Sexualized Violence Moderated by Rape Myth Acceptance and Benevolent Sexism. Violence Against Women.

Halicki, K. T., Suitner, C., Vogel, T., & Wänke, M. (2021). What is agentic about the Spatial Agency Bias? How pragmatic relevance contributes to the spatial representations of actions. European Journal of Social Psychology, 51(1), 166-177.

Hegarty, P., & Buechel, C. (2006). Androcentric reporting of gender differences in APA journals: 1965–2004. Review of General Psychology, 10(4), 377-389.

Kesebir, S. (2017). Word order denotes relevance differences: The case of conjoined phrases with lexical gender. Journal of personality and social psychology, 113(2), 262.

Humphrey, N. K., & McManus, I. C. (1973). Status and the left cheek. New Scientist, 59, 437-439.

Maass, A., & Russo, A. (2003). Directional bias in the mental representation of spatial events: Nature or culture? Psychological Science, 14(4), 296-301.

Maass, A., Pagani, D., & Berta, E. (2007). How beautiful is the goal and how violent is the fistfight? Spatial bias in the interpretation of human behavior. Social Cognition, 26(6), 833-852.

Maass, A., Suitner, C., & Deconchy, J. P. (2014). Living in an asymmetrical world: How writing direction affects thought and action. Psychology Press.

Maass, A., Suitner, C., & Nadhmi, F. (2014). What drives the spatial agency bias? An Italian–Malagasy–Arabic comparison study. Journal of Experimental Psychology: General, 143(3), 991.

Maass, A., Suitner, C., Favaretto, X., & Cignacchi, M. (2009). Groups in space: Stereotypes and the spatial agency bias. Journal of Experimental Social Psychology, 45(3), 496-504.

Mazzurega, M., Paladino, M. P., Bonfiglioli, C., & Timeo, S. (2019). Not the right profile: Women facing rightward elicit responses in defence of gender stereotypes. Psicologia sociale, 14(1), 57-72.

McManus, I. C. (2005). Symmetry and asymmetry in aesthetics and the arts. European Review, 13(S2), 157-180.

Mendonça, R., Garrido, M. V., & Semin, G. R. (2020a). Asymmetric

practices of reading and writing shape visuospatial attention and discrimination. Scientific Reports, 10, 21100. https://doi.org/10.1038/s41598-020-78080-0

Mendonça, R., Garrido, M. V., & Semin, G. R. (2020b). Social inferences from faces as a function of the left-to-right movement continuum. Frontiers in psychology, 11, 1488.

Mendonça, R., Garrido, M. V., & Semin, G. R. (2022). Two Cultural Processing Asymmetries Drive Spatial Attention. Cognitive Science, 46(8), e13185.

Monahan, L., & Romero, M. (2020). Heading the right way? The influence of motion direction in advertising on brand trust. Journal of Advertising, 49(3), 250-269.

New, J., Cosmides, L., & Tooby, J. (2007). Category-specific attention for animals reflects ancestral priorities, not expertise. Proceedings of the National Academy of Sciences of the United States of America, 104(42), 16598–16603. https://doi.org/10.1073/pnas.0703913104

Nicholls, M. E., Clode, D., Wood, S. J., & Wood, A. G. (1999). Laterality of expression in portraiture: Putting your best cheek forward. Proceedings of the Royal Society of London. Series B: Biological Sciences, 266(1428), 1517-1522.

Pérez González, C. (2012). Lateral organisation in nineteenth-century studio photographs is influenced by the direction of writing: A comparison of Iranian and Spanish photographs. Laterality: Asymmetries of Body, Brain and Cognition, 17(5), 515-532.

Payne, D. L. (Ed.). (1992). Pragmatics of word order flexibility (Vol. 22). Amsterdam: John Benjamins.

Sebanz, N., Bekkering, H., & Knoblich, G. (2006). Joint action: bodies and minds moving together. Trends in Cognitive Sciences, 10(2), 70–76. https://doi.org/10.1016/j.tics.2005.12.009

Siewierska, A., Rijkhoff, J., & Bakker, D. (2010). Appendix–12 word order variables in the languages of Europe. In Siewierska, A. (Ed.), Constituent order in the languages of Europe (Vol. 20, No. 1, pp. 783–811). Berlin: de Gruyter.

Singh, M., Vaid, J., & Sakhuja, T. (1987). Directional scaning and handedness: Influence on the direction and orientation of free-hand drawing. Poster presented at nuaj çonference of the National Academy of Neuropsychology, Chicago.

Spalek, T. M., & Hammad, S. (2005). The left-to-right bias in inhibition of return is due to the direction of reading. Psychological Science, 16(1), 15-18.

Spence, J. T., & Helmreich, R. L. (1978). Masculinity and femininity: Their psychological dimen- sions, correlates, and antecedents. Austin: University of Texas Press.

Suitner, C., & Maass, A. (2007). Positioning bias in portraits and self-portraits: Do women make different decisions? Empirical Studies of the Arts, 25, 71–95.

Suitner, C., & Maass, A. (2008). The role of valence in the perception of agency and com- munion. European Journal of Social Psychology, 38, 1073–1082.

Suitner, C., & Maass, A. (2016). Spatial Agency Bias: Representing People in Space. Advances in Experimental Social Psychology, 53, 245-301.

Suitner, C., Maass, A., & Ronconi, L. (2017). From spatial to social asymmetry: spontaneous and conditioned associations of gender and space. Psychology of Women Quarterly, 41(1), 46-64.

Suitner, C., Maass, A., Bettinsoli, M. L., Carraro, L., & Kumar, S. (2017). Left-handers' struggle in a rightward wor (l) d: The relation between horizontal spatial bias and effort in directed movements. Laterality: Asymmetries of Body, Brain and Cognition, 22(1), 60-89.

Suitner, C., Maass, A., Navarrete, E., Formanowicz, M., Bratanova, B., Cervone, C., Hakoköngäs J. E., Kuppens, T., Lipourli, E., Rakić, T., Scatolon, A., Teixeira, C. P., Wang, Z., Sobral, M. P., & Carrier, A. (2021). Spatial agency bias and word order flexibility: A comparison of 14 European languages. Applied Psycholinguistics, 42(3), 657-671.

Tajfel, H., & Turner, J. C. (1982). Social psychology of intergroup relations. Annual review of psychology, 33(1), 1-39.

ten Cate, C. (2002). Posing as a professor: Laterality in posing orientation for portraits of scientists. Journal of Nonverbal Behavior, 26(3), 175-192.

von Hecker, U., & Klauer, K. C. (2021). Spatial processes in category assignment. Journal of Experimental Psychology: General, 150(3), 446.

von Hecker, U., Arjmandi Lari, Z., Fazilat-Pour, M., & Krumpholtz, L. (2022). Attribution of feature magnitudes is influenced by trained reading-writing direction. Journal of Cognitive Psychology, 34(2), 194-207.

Weis, P. P., Nikadon, J., Herbert, C., & Formanowicz, M. (2022). The verb–self link: An implicit association test study. Psychonomic Bulletin & Review. https://doi.org/10.3758/s13423-022-02105-0

Wilson, M. (2002). Six views of embodied cognition. Psychonomic Bulletin & Review, 9(4), 625-636.

Zebian, S. (2005). Linkages between number concepts, spatial thinking, and directionality of writing: The SNARC effect and the reverse SNARC effect in English and Arabic monoliterates, biliterates, and illiterate Arabic speakers. Journal of Cognition and Culture, 5(1-2), 165-190.

10. The Everyday Diagrams of Social Relationships: Drawing Gender, Kinship, Work, and Sexuality Together with the Spatial Agency Bias

Peter Hegarty
The Open University, UK

It is a privilege and a pleasure to be invited to this academic family gathering honouring Anne Maass and her influence. Ironically enough, my contribution will examine how families are conceptualized and visualized. Here, I extend my interest in Anne's work on the Spatial Agency Bias (SAB) from bar graphs of gender differences to other everyday diagrams; kinship diagrams and organizational charts. Some new studies are presented and used as an occasion to imagine the relevance of Anne's rediscovery of the SAB to feminist studies beyond psychology.

The Spatial Agency Bias (SAB) and Graphs of Gender Differences

The SAB is a hypothesized bias affecting how we encode, imagine and depict social agency in the direction of the language that we learn to read and write. In Italian, for example, this direction is left-to-right and in Arabic it is right-to-left. The SAB hypothesis is less relevant to languages written in vertical directions or written in multiple directions. In regard to gender stereotypes, the bias predicts that men would be depicted *first* before women in the horizontal writing direction; men go left of women in Italian and right of women in Arabic (Suitner & Formanowicz, this volume; Suitner & Maass, 2016).

Consistent with the SAH hypothesis, people array data representing males first before females in graphs and tables.[1] For example, in psychology articles published 1965-2004, about three quarters of the tables and graphs communicating gender differences array data representing males first with regard to writing direction (Hegarty & Buechel, 2006). This preference can be found in medical articles reporting gender differences also (Rudin et al., 2009). Recent editions of the American Psychological Association's publication manual have cited this finding to urge authors to be mindful that graph and table order may implicitly communicate which group is the universal standard (American Psychological Association, 2010, 2020). Subsequent experiments also found a preference to graph data representing males ahead of data representing females, and that gender difference graphs are selectively mis-remembered by flipping female-first order to male-first order. The graph order preference emerges not because men are the universal standard. Rather, as per the SAB hypothesis, men are perceived as more agentic than women. Indeed people graphing in English spontaneously graph more powerful groups left and weaker groups to the right (Hegarty et al., 2010, Studies 3-4: Hegarty & Parr, in press).

The graph order preference relates the SAB hypothesis to psychologists' own thinking. It compels us to recognize that psychology is a *reflexive* science; the psychologists who produce it can also be the objects of its theories (Morawski, 1994; Richards & Stenner, 2022). Reflexive human sciences are characterized by "feedback loops" between the people who are objects of those sciences and the people who theorize about them, making them inherently more historical and cultural endeavours than the natural sciences mimicked by our experimental methods (see Hacking, 1995). Whilst the APA Publication Manual has long promoted a natural science epistemology, its later editions also recognize that real people, by virtue of their social identities, are impacted by the science written about them (Sigal & Pettit, 2012). The evolution of the manual's advice about gender, including early arguments against male generics and current advice on recognizing non-binary gender evidences a history of feedback loops between research on scientific representation and social norms that organize how authors represent their research participants to their scientific peers (see Roessel & Merkel, this volume).

---

1 I refer to the groups represented in graphs and tables as "females" and "males" rather than "women" and "men" to be inclusive of the many representations of human children described by psychologists.

Such reflexivity makes the graph order effect and the SAB relevant to questions in feminist philosophy. For decades, three different feminist epistemologies to science have been described as *empiricism feminism, standpoint feminism,* and *feminist postmodernism* (see Harding, 1986, and Riger, 1992 for an application to psychology). In psychology, empiricism feminism amounts to standard use of quantitative methods to examine gender differences and gender stereotypes. Standpoint feminism acknowledges that women and men are co-created through relationships of dominance and subordination, and that women, like other subordinate groups, can have an "epistemic advantage" over men in the way that they see the world. Finally, postmodern feminist tends to draw on radical relativist epistemologies to argue that psychological science, like any other science, can never be firmly grounded.

In this chapter, I want to present some new data extending research on the graph order effect to a new format *and* to comment on how the effect informs the productive tensions between different strands of feminist epistemology. First, psychologists Alice Eagly and Stephanie Riger (2014) have shown that in the last half century feminism has influenced psychology by diversifying the genders of its authors and study participants but has not changed reliance on experimentation and quantification. In other words, feminist empiricism has been very successful. These authors described the graph order effect as a subtle scientific practice that continues unabated in this empirical feminist mainstream. Second, Hegarty and Buechel (2006) and Hegarty et al. (2010) observed that both women and men graphed gender differences with data representing men prioritized to the left of data representing women. Philosopher Janet Kourany (2010) described the graph order effect as illustrating how women scientists can routinely produce androcentric science, at odds with the assumption of standpoint theorists that women routinely have an "epistemic advantage" in regards to scientific understanding and representation of gender-related matters. Finally, graphs have been central to relativist and post-modernist arguments about science. Perhaps the strongest expression of this epistemology is the graphism thesis is that expressed by Bruno Latour (1990), that by drawing things together graphs can create the illusion that one is looking at "the thing itself" rather than a human-made representation (see Smith et al., 2002 for an application to psychology). The preference to graph data representing males first does not support radical relativist arguments directly. Rather, by making the graphism question more

relevant to gender, the psychological work that I have conducted with colleagues calls attention to some unexamined androcentric metaphors within Latour's argument for the graphism thesis (Hegarty & Lemieux, 2011).

The APA publication manual's citation of the graph order preference also makes the SAB relevant to the question of how scientific practices change in response to new norms, a pressing empirical question in the context of psychology's rapidly shifting relationship with open science practices. To explore this question, Amy Parr and I conducted two content analyses replicating and extending Hegarty and Buechel's original study to the decade since 2010 when the 6th edition of the APA publication manual first drew authors' attention to graph and table order as a possible source of bias (Hegarty & Parr, 2024). We found that in developmental psychology, the traditional males-first preference remains. However, in social psychology, men authors graph males first even more than before whilst women authors graph females and males first with approximate equal frequency. These shifts in representational practices recall shifts in the arts in an early century. In the 19th century, women artists changed the direction of sitters' postures to oppose a convention that represented men as more agentic than women (Suitner & Maass, 2016). Women social psychologists seem to similarly be changing the visualization of gender differences in social psychology now.

Word Order Preferences: Masculine-First and Closer-First

Anne has long taught social psychologists that subtle linguistic choices carry meaning (Maass, 1989). The SAB is influenced not only by writing direction but also by the order in which the agents and objects of verbs are named in the active voice in one's language. Most human languages place the subject before the object of the verb in the active voice. Those rare languages, such as Malagasy, that do the opposite, seem to prompt imagery in which the object precedes the subject in space (Maass et al., 2014).

Somewhat downstream in the language production process sits the ordering of lexically equivalent terms in binomial phrases (e.g., "gin and tonic," "here and now" and "men and women"). Linguists have long thought of binomial phrases as having relatively "frozen" or fixed word order (Malkiel, 1959). The linguist Sandra Mollin (2014) has shown that over historical time, many binomials "unfreeze" so that preferences for

order weaken or reverse. For example, the behavioural preference to name men first in many binomial phrases for kinfolk (e.g., uncle and aunt, brother and sister) unfroze in the late 20th century at the same time that feminists challenged many other aspects of sexist language use (Cameron, 1998).

In one line of studies, my colleagues and I investigated the naming of specific couples within binomial phrases (e.g., "William and Kate"). Previous authors had argued that differences in the phonology, frequency and length of women's and men's names created a "conspiracy" to name male partners of couples first (Wright et al., 2005). We argued that gender stereotypes create a masculine-first preference (Hegarty et al., 2011). Gender stereotyping doesn't occur when we have individuating information about others such as our friends and family (Fiske & Neuberg, 1989). When naming imaginary couples, participants lack such information and named the more stereotypical masculine partner in imaginary couples first. I use the term "masculine" here as participants in one study also named the more stereotypically masculine partner of same-sex couples first (Hegarty et al., 2011, Study 4). However, we are less likely to stereotype friends and family, and name the closer person first, not the more stereotypically masculine (Hegarty et al., 2011, Study 5).

Graphs and names get drawn together in a variety of everyday diagrams; kinship diagrams and organizational charts. Does the SAB have influence when women and men are drawn into these representations of social relationships? How might the rules of order be different when we imagine women and men in diverse familial and occupational roles?

Kinship Diagrams: How Do Gender Stereotypes Run in the Family?

To develop the relevance of the graph order effect to feminist epistemology, I turned to examine its impact on kinship diagrams. Family is not studied as much as some other bases of social identity in social psychology. However, kinship is central to cultural anthropology. David Schneider (1968) critiqued previous studies of kinship in anthropology in a reflexive and relativist way. Whilst anthropologists had often studied cultural variation in kinship systems, they had acted as if certain natural facts of kinship were universal among human cultures. Schneider (1968) studied White middle-class Americans, finding their constructions of family and kinship to be organized by aligning two elements: shared bio-genetic substance (or "blood") and a diffuse enduring solidarity ("family ties"). He

further argued that anthropologists had imposed these cultural constructions on other cultures, treating those cultures' schemes as needing explanation to the extent that they deviated from these seemingly "natural" but really "American-cultural" facts. Schneider's work is exemplary of a mode of empirical critique that led a discipline to see how it had imposed a cultural world view as a standard on others in the name of seemingly-objective theory. Of course, this is precisely the kind of problem that seemingly-objective male-centred science introduces, and that all forms of feminist epistemology aim to resist. Schneider concluded that there were in fact no natural facts about kinship at all (see Stone 2004, for a cogent summary and critique).

Kinship diagrams were not central to Schneider's critique, but they have long been central to anthropologists' ways of objectifying others' cultural understandings of kinship. Bouquet (1996) traced the origins of 20th century anthropologists' kinship diagrams back through to the tree imagery common in 18th and 19th century Europe sciences to the Christian Bible. The family tree diagram form is not only ancient, it is also popular in modern times. The corporation ancestry.com boasts that its members have created "over 100 million family trees on Ancestry ® most of which are public." (Ancestry, com, n.d.). If culture "consists in the way people draw analogies between different domains of their worlds" (Strathern, 1992, p. 47), then testing the SAB hypothesis in the domain of family trees allows us to assess whether family ties between women and men are visualized according to an embodied bias that prioritizes men.

Three experiments, all conducted with English language speaking students in the UK, tested the SAB hypothesis that men would be arrayed first (i.e., on the left) when drawing the trees of imagined families. Studies 1a and b presented participants with a diagram representing three-generations of a family including two pairs of grandparents, one pair of parents and three siblings (see Table 1). Study 1c presented a four-generation family that also included four sets of great-grandparents, but only two siblings. In all studies, participants were asked to write names into the boxes on the trees to represent a typical but imaginary family. Studies 1a and 1b were conducted at university open days in a taster session on psychological research. Study 1c was conducted by Gledis Shahu (2018) as part of her M.Sc. dissertation at the University of Surrey.

The number of students in each study, the proportion who were women, and the proportion of cases confirming the SAB hypothesis in each experiment are shown in Table 1.

Table 1: Participant Characteristics and the Percentage of Couples Arrayed with Men First and of Siblings Named Male (Studies 1a, b, c).

Study	1a	1b	1c	All
Participants				
N	81	59	181	321
Proportion Women	83%	49%	75%	72%
Proportion with Men on Left				
Great-Grandparents (n=4)	---	---	66%	66%
Grandparents (n = 2)	62%	78%	58%	63%
Parents (n = 1)	55%	68%	59%	60%
Proportion Named as Men				
Leftmost Sibling	52%	69%	44%	50%
Middle Sibling	54%	57%	---	56%
Rightmost Sibling	55%	48%	56%	54%

Table 1. Table with Participant Characteristics and the Percentage of Couples Arrayed with Men First and Siblings Named Male (Studies 1a, b, c).

Several patterns are clear. First, when arranging the names of heterosexual couples, the SAB hypothesis was confirmed; most often men's names were positioned first, on the left, in each experiment. Indeed, support for the SAB was somewhat stronger among couples in older generations than in younger generations. This finding recalls how participants named imaginary heterosexual couples with men's names first more often if the couples were described as marrying a long time ago rather than in recent decades (Hegarty et al., 2011, Study 3).

Second, participants more often gave the siblings male names than female names, but did so irrespective of their position. In Study 1a, we also asked participants to note each sibling's age. Participants were arrayed in decreasing age from left (M age = 12.85), to centre (M age = 11.46 years), to right (M age = 10.7 years), within-subjects $F_{(1,79)} = 7.77$, $p = .007$, $\eta p2 = .09$. There was an order convention here, but it was based on

age and not gender, and is consistent with the implicit representation of time moving in the direction of writing.

Finally, Studies 1b and 1c included experimental manipulations which had no effect on the support for the SAB.   In Study 1b, participants were randomly assigned either to think about a traditional or non-traditional family. This manipulation had no effect, in contrast to an earlier finding that it impacted the ordering of imaginary couples names outside of a diagrammatic context (Hegarty et al., 2011, Study 2). Study 3b presented an abstract kinship diagram or one overlaid with the concrete image of a tree. This manipulation had no effect.

Organizational Charts: Do We Take the SAB to Work

Tree diagrams are not only used to represent domestic relationships they are also part of the culture at work. Organizational charts, also later called "organigrams" or "organograms" date from the 19th century, but did not become popular until the middle of the 20th century. These visual metaphors represent the structure of organizations broken down into specific jobs and roles, often naming the individual role-holders. Organizational charts, like kinship diagrams, use spatial extension and lines connected people's names to represent social relationships between individuals. The ways that they are drawn have previously been shown to express embodied spatial biases. For example, people elongate the vertical dimension of organizational diagrams to the extent that they perceive the power difference between organizational brackets to be large (Giessner & Schubert, 2007).

Study 2 tested the SAB hypothesis in regard to organizational charts. One hundred and fifty-five women and 136 men took part by adding names to a blank organizational chart including one manager who supervised two middle-managers, each of whom supervised two workers. The participant volunteers completed the materials as a distractor task in an unrelated experiment on person memory for a class project. Each name was coded as 0 (female) or 1 (male).   I also tested an SAB   hypothesis among the four workers; that leftward individuals would be attributed more power than rightward individuals. Participants ranked how much power was possessed by each of the four workers from most (1) to least (4). These rankings were reverse coded from 1-4 so that higher numbers reflected greater power.

Participants' imagination of this workplace was clearly androcentric (see Bailey et al., 2019), they gave far more employees men's names than women's names. Men's names were given to 84% of senior managers, 68% of middle-managers, and 67% of workers. This finding is consistent with findings that workers, particularly successful managers, are assumed to be men by default (e.g., Schein et al., 1996).

Overall, the SAB hypothesis received little support. Among the middle-managers, the ratio of women's to men's names was the same for the leftward manager (33% men vs 67% women) and the rightward manager (31% men vs 69% women). Second, one-way within subjects ANOVA showed that there was no significant decrease in the proportion of men placed in the four worker positions moving from left to right within the organizational chart (72%, 59%. 60%, 62% respectively), $F (1,290) = 1.64$, $p = .20$, $\eta_p^2 = .006$. Finally, we tested whether workers on the left of the diagram would be accorded more power than workers on the right. There was only a marginally significant difference between workers by position from left to right (M_{Power} = 2.55, 2.61, 2.47, 2.36 respectively), $F(1,273) = 3.81$, $p = .052$, $\eta_p^2 = .01$. The results and implications of all studies are discussed next.

The SAB and Feminist Studies: Denaturalizing Kinship and Heterosexuality

These studies tested the SAB hypothesis in context where named individuals are drawn together with their familial or occupational social relationships represented by lines and special arrangements. Support for the SAB hypothesis was strongest when naming heterosexual couples in kinship diagrams, rather than siblings in those diagrams, or co-workers in organizational charts. Clearly gender stereotypes created preferences to order names from left-to-right men first in some cases but not others. How should we make sense of this qualified support for the SAB hypothesis here?

First, this pattern of data can't be easily explained by general differences between women's and men's names in length, phonology or frequency (e.g., Wright et al., 2005), because the SAB was sometimes supported but not always. By the same logic, the data can't simply be an effect of cognitive androcentrism, the tendency to draw men to mind more readily than women (Bailey et al., 2019). Indeed, support for the SAB was completely absent in Study 2, even though participants overwhelmingly called men

workers to mind here. Men may have been considered more relevant to the work context, but they were not positioned first consistently (see Kesebir, 2017). Only when androcentrism was constrained by the implicit demand to make up couples with one woman's name and one man's name did strong evidence of the SAB emerge (Study 1).

How might coupling people up and locating them as the parents of children also gender them? Building on Schneider's relativist critique, anthropologists Sylvia Yanagisako and Jane Collier argued that (hetero) sexual reproduction was central not only to American understandings of kinship, but also to a particular American understanding of the biological distinction between women and men. They argued that studies of kinship which impressed the American understanding of kinship on other cultures also projected the assumption that the difference between women and men was ultimately grounded in differences in women's and men's roles in biological reproduction. In other words, unreflexive thinking about kinship in anthropology had "made people up" as two discrete classes of women and men. Perhaps thinking about people as couples situated in kinship structures activates certain gender concepts that thinking about people as siblings or as co-workers does not.

Yanigisako and Collier's (1987) feminist anthropological critique is resonant with arguments which soon emerged in queer theory that people are not naturally members of the categories "women" and "men," but become intelligible as such by virtue of dominant cultural heteronormative ideologies that conflate "true" or "real" gender with heterosexual gender (e.g., Butler, 1990, see Morris, 1993 for an integration of early queer theory and anthropology). By de-naturalizing the reproductive heterosexual couple, Schneider opened up new anthropological studies of gay/lesbian kinship. American lesbian women and gay men learned the hard way that "family ties" were not natural at all when they came out to kinfolk who rejected them. American lesbians and gay men also made up kinship ties differently using artificial insemination during the HIV pandemic, when blood and semen took on new cultural significance in signifying social familial relationships (see Weston, 1993).

The methods and results of Study 1 operated entirely within a heteronormative frame. Participants were not explicitly instructed to call to mind heterosexual couples in these families, but with vanishingly rare exceptions, they always gave the two individuals in the couple's names common to different genders. Yanagisako and Collier (1987) were explicit that the genres of representation in kinship diagrams played a constitu-

tive part of this process of making people up as members of two distinct gender groups:

> The standard units of our genealogies, after all, are circles and triangles about which we assume a number of things. Above all, we take for granted that they represent two naturally different categories of people that the natural difference between them is the basis of human reproduction and, therefore, kinship.

As such, these authors make the point that kinship diagrams are not merely empty conventions (see also Bouquet, 1996), but meaning systems that gender people within a binary logic where difference is defined by reproductive sex. A kinship diagram with only one shape or a myriad number of shapes might seem queer in comparison.

In conclusion, I think that Anne Maass' re-discovery of the SAB makes a unique contribution to this kind of reflexive feminist analysis of dominant systems of meaning-making about gender in the man-made human sciences not only by offering a psychological level of explanation, but in drawing in embodiment. The philosopher Elizabeth Grosz (1994) has persuasively argued that feminism requires an account of embodiment that deconstructs a long-standing Western philosophical tradition that separates disembodied reason (gendered male) from the embodiment that is projected onto women who are consequently deemed incapable of reason (see also Rooney, 1991). Describing gender differences in women's and men's thinking - as standpoint feminist theories can do - does not necessarily provide such a useful feminist analysis. Standpoint theories can risk essentializing, polarizing and stereotyping differences between women's and men's knowledge, understanding or thinking (Riger, 1992). Consequently, sociologists have argued feminism requires a theory of how knowledge is embodied that does not simply theorize gender in terms of biological sex differences taken out of social and cultural contexts (see Bluhm, Jacobson, & Maibom, 2012; Longino, 2010; Pitts-Taylor, 2015 for discussion). In other words, feminist theories of embodied knowledge need to understand how knowledge is embodied without reifying the particular nexus of cultural assumptions about kinship, gender, and reproduction that Yanagisako and Collier (1987) called on feminists to expose and deconstruct.

The SAB seems to exemplify the kind of theory of embodied cognition that feminism needs. I take seriously the claims that graphs and diagrams are visual metaphors that use space to represent other properties (Latour, 1990), and that culture is about the way in which we draw

analogies between different realms of experience (Strathern, 1992). Consequently, the SAB has clear application in explaining how metaphorical graphical representations – including those used in science – are gendered forms of culture, that are not the "way of knowing" of any one sex or gender group but become embodied by the way we learn to write and read. In other words, I think that the SAB has particular inter-disciplinary relevance for feminist anthropology, sociology and philosophy because it is an enculturated embodied bias with unique power to explain how gendered representations can become the default for artistic or scientific genres of representation in literate cultures. If we have made ourselves up as Anne Maass' academic family in this volume, then I humbly suggest that her work in this area is a means by which we might make up new relationships with our cousins in neighbouring disciplines.

References

American Psychological Association. (2010). Publication manual of the American
Psychological Association, (6th Ed.). American Psychological Association.
American Psychological Association. (2020). Publication manual of the American
Psychological Association, (7th Ed.). American Psychological Association.
Ancestry.com (n.d.). Searching public family trees.
https://support.ancestry.co.uk/s/article/Searching-Public-Family-Trees. Downloaded 27th Sept, 2022.
Bailey, A.H., LaFrance, M., & Dovidio, J. F. (2019). Is man the measure of all things: A social cognitive account of androcentrism. Personality and Social Psychology Bulletin, 23 (4), 303-317.
Bluhm, R., Jacbson, A.J., & Maibom, H.L. (2012). Introduction. In R. Bluhm, A.J. Jacobson, & H.L. Maibom (Eds)., Neurofeminism: Issues at the intersection of feminist theory and cognitive science. Palgrave Macmillan.
Bouquet, M. (1996). Family trees and their affinities: the visual imperative of the genealogical diagram. Journal of the Royal Anthropological Institute, 43-66.
Butler, J. (2002). Gender trouble: Feminism and the subversion of identity. Routledge.
Cameron, D. (Ed.). (1998). The feminist critique of language: A reader. Routledge.

Eagly, A. H., & Riger, S. (2014). Feminism and psychology: Critiques of methods and epistemology. American Psychologist, 69(7), 685-702.

Fiske, S. T., & Neuberg, S. L. (1990). A continuum of impression formation, from category-based to individuating processes: Influences of information and motivation on attention and interpretation. In Advances in experimental social psychology (Vol. 23, pp. 1-74). Academic Press.

Giessner, S. R., & Schubert, T. W. (2007). High in the hierarchy: How vertical location and judgments of leaders' power are interrelated. Organizational Behavior and Human Decision Processes, 104, 30-44.

Grosz, E. (2020). Volatile bodies: Toward a corporeal feminism. Routledge.

Hacking, I. (1995). The looping effects of human kinds. In D. Sperber, D. Premack, & A. J. Premack (Eds.), Causal cognition: A multidisciplinary debate (pp. 351–394). Oxford University Press.

Harding, S. G. (1986). The science question in feminism. Cornell University Press.

Hegarty, P., & Buechel, C. (2006). Androcentric reporting of gender differences in APA journals: 1965-2004. Review of General Psychology, 10 (4), 377-389.

Hegarty, P., & Lemieux, A. (2011). Who is the second (graphed) sex and why? The meaning of order in graphs of gender differences. In T. W. Schubert. & A. Maass (Eds.), Spatial Dimensions of Social Thought (pp. 325-349). Mouton de Gruyter: Berlin.

Hegarty, P., Lemieux, A. F., & McQueen, G. (2010). Graphing the order of the sexes: Constructing, recalling, interpreting, and putting the self in gender difference graphs. Journal of Personality and Social Psychology, 98 (3), 375-391.

Hegarty, P., Watson, N., Fletcher, K., & McQueen, G. (2011). When gentlemen are first and ladies are last. Effects of gender stereotypes on the order of romantic partners' names. British Journal of Social Psychology, 50, 21-35.

Hegarty, P., & Parr, A. (2022). Embodied standpoints in gender differences graphs and tables: When, where and why are men still prioritized? Manuscript Under Review.

Kesebir, S. (2017). Word order denotes relevance differences: The case of conjoined phrases with lexical gender. Journal of Personality and Social Psychology, 113(2), 262–279.

Latour, B. (1990). Drawing things together. In M. Lynch & S. Woolgar (Eds.), Representation in scientific practice (pp. 19–68). MIT Press.

Longino, H. (2010). Feminist epistemology at Hypatia's 25th anniversary. Hypatia,25(4), 733-741.

Kourany, J. A. (2010). Philosophy of science after feminism. Oxford University Press.

Maass, A. (1999). Linguistic intergroup bias: Stereotype perpetuation through language. In Advances in experimental social psychology (Vol. 31, pp. 79-121). Academic Press.

Maass, A. (1999). Linguistic intergroup bias: Stereotype perpetuation through language. In Advances in Experimental Social Psychology (Vol. 31, pp. 79-121). Academic Press.

Maass, A., Suitner, C., & Nadhmi, F. (2014). What drives the spatial agency bias? An Italian–Malagasy–Arabic comparison study. Journal of Experimental Psychology: General, 143(3), 991-996.

Malkiel, Y. (1959). Studies in irreversible binomials. Lingua, 8, 113-160.

Mollin, S. (2014). The (ir) reversibility of English binomials: Corpus, constraints, developments. John Benjamins.

Morawski, J. G. (1994). Practicing feminisms, reconstructing psychology: Notes on a liminal science. University of Michigan Press.

Morris, R. C. (1995). All made up: Performance theory and the new anthropology of sex and gender. Annual Review of Anthropology, 567-592.

Pitts-Taylor, V. (2015). A feminist carnal sociology? Embodiment in sociology, feminism, and naturalized philosophy. Qualitative Sociology, 38 (1), 19-25.

Richards, G., & Stenner, P. (2022). Putting psychology in its place: Critical historical perspectives (4th edition). Routledge.

Riger, S. (1992). Epistemological debates, feminist values: Science, social values, and the study of women. American Psychologist, 47 (6), 730-740.

Rooney, P. (1991). Sex metaphor and conceptions of reason. Hypatia, 6 (2), 77-103.

Rudin, S.I., Jones, M.O., Lemieux, A.F., & Hegarty, P. (2009). Androcentrism in medical journals: The communication of bias through visuospatial displays. Poster presented at the Annual Purchase College Natural and Social Sciences Student Research Symposium.

Schein, V. E., Mueller, R., Lituchy, T., & Liu, J. (1996). Think manager—think male: A global phenomenon?. Journal of Organizational Behavior, 17(1), 33-41.

Schneider, D. (1968). American kinship: A cultural account. University of Chicago Press.

Shahu, G. (2018). The influence of gender stereotypes on the drawing of family trees # Unpublished MSc dissertation: University of Surrey, UK.

Sigal, M. J., & Pettit, M. (2012). Information overload, professionalization, and the origins of the publication manual of the American Psychological Association. Review of General Psychology, 16(4), 357-363.

Smith, L. D., Best, L. A., Stubbs, A. D., Archibald, A. B., & Roberson-Nay, R. (2002). Constructing knowledge: The role of graphs and tables in hard and soft psychology. American Psychologist, 57(10), 749–761.

Stone, L. (2004). The demise and revival of kinship: Introduction. In R. Parkin and L. Stone (Eds)., Knship and family: An anthropological reader (pp. 241-256). Blackwell.

Strathern, M. (1992). Reproducing the future: Anthropology, kinship and the new

reproductive technologies. New York: Routledge.

Suitner, C., & Formanowicz, M. Spatial agency bias; Mapping social agency into the visual field. (This volume).

Suitner, C., & Maass, A. (2016). Spatial agency bias: Representing people in space. In J. M. Olson & M. P. Zanna (Eds.), Advances in Experimental Social Psychology (Vol. 53, pp. 245-301). Elsevier Academic Press.

Suitner, C., & Maass, A. (2007). Positioning bias in portraits and self-portraits: Do female artists make different choices?. Empirical Studies of the Arts, 25(1), 71-95.

Weston, K. (1997). Families we choose: Lesbians, gays, kinship. Columbia University Press.

Wright, S., Hay, J., & Bent, T. (2005). Ladies first? Phonology, frequency, and the naming conspiracy. Linguistics, 43, 531–561.

Yanagisako, S.J., & Collier, J.F. (1987) toward a unified analysis of gender and kinship. In S.J. Yanagisako & J.F. Collier (Eds)., Gender and kinship: Essays toward a unified analysis (pp. 14-50). Stanford University Press.

11. On Space, Speed and Abstraction

Nira Liberman[1], Ayelet Hatzek[2], Yaacov Trope[3], and Ravit Nussinson[2]
[1]Tel Aviv University, Israel
[2]The Open University of Israel
[3]New York University, New York, USA

The work by Anne Maass has taught us much about how language not only reflects but also influences the way we perceive individuals, groups, and their interrelations. For example, how we choose to describe an action affects ascription of agency and legal responsibility (Fausey & Boroditsky, 2010; Formanowicz et al., 2017), and is affected by whether the action is positive or negative in combination with whether it was performed by a member of the ingroup or an outgroup (Maass et al., 1989; 1996). Anne Maass' work also tells us that effects of language extend even to perceptions of space and spatial relations. In particular, writing direction (left-to-right, as in English and other European languages, versus right-to-left, as in Arabic and Hebrew) affects whether we ascribe power, speed, and agency to a person or a moving object that are located on the left versus the right side.

There are multiple points of interface between Anne Maass' seminal work and Construal Level Theory (CLT, Liberman, et al., 2007; Liberman & Trope, 2008; 2014; Trope & Liberman 2010), perhaps the most important being that both share the same fundamental view that people function in a grid of space, time, social relations, and possibilities and that language is (part of) the medium that connects entities on that grid.

In CLT, different dimensions of distance share important commonalities (a common zero point, a common way to transcend them via using higher-level, more abstract mental and social tools), tend to co-occur in people's everyday experiences (e.g., it typically takes more time to go to a

farther away place), and therefore are interrelated. As a result, correlates of one distance tend to also appear with other distances. Moreover, correlates of farther distance may both reflect and regulate distance on all dimensions. For example, addressing a person in a polite, formal way (in contrast to a colloquial way) is designed to both signal and create social distance. Consistent with the framework of CLT, it does so by using abstract language (e.g., euphemisms), spatial distance (e.g., farther standing distance), hypotheticality (avoiding assertions), and temporal distancing (e.g., preferring future to present tense; Stephan et al., 2010). Moreover, any of these (e.g., farther spatial distance; using more abstract language) can both express and create social distance versus closeness (Stephan et al., 2011).

The work on Spatial Agency Bias (SAB, Suitner & Maass, 2016) uncovered important regularities that come under this general umbrella. This line of work has shown that correlates of spatial distance, in this case the spatial organization of written text, extend to social distance and to temporal distance. For example, in rightward-writing languages, the left-end of a horizontal line is perceived as the beginning, and thus as closer to me, here and now. Studies conducted within this framework indeed showed that in such languages the self is positioned to the left of the other, the ingroup to the left of the outgroup, and also that the left occurs before the right (e.g., running from left to right is perceived as more natural and as being faster, more efficient, than running from right to left, reference). Consistent with SAB, many of these phenomena were shown to reverse in cultures with a right-to-left writing direction (Maass et al., 2014). In SAB, the aforementioned findings are explained by agency: The initiator of an action, the self and the ingroup are ascribed more agency, and to the extent that this occurs they are placed where writing begins. The perspective of CLT is subtly different. Namely, because distances are interrelated, the proximal ends of each dimension (near me in space, earlier in time, closer to me on the social dimension) are placed together at the spatially proximal end of a line, which is determined by the direction of writing.

It would be interesting to examine whether, as CLT would predict, a similar regularity would emerge also with the dimension of hypotheticality. For example, would a picture that is placed on the left side seem to rightward-writing participants more real (as opposed to imaginary/hypothetical), and as depicting a more likely situation than a similar picture that is placed on the right? An extension of this logic to construal level

is also potentially interesting: Would actions, events and objects that are depicted on the left (vs. the right) side on a horizontal line be construed more concretely in rightward writing cultures? Would this relation reverse in leftward-writing cultures? Relatedly, would concrete words (e.g., bread) and abstract words (e.g., custom) be easier to process when the concrete word appears to the left of the abstract word than vice versa in rightward-writing cultures (Amit et al., 2019)?

The relations between SAB and CLT run deep, and acknowledging them has the potential to elucidate research in both frameworks. We now turn to examine from the perspective of SAB two lines of research that have been conducted within the framework of CLT, and which connected speed and vertical position to psychological distance and abstraction.

Speed

In SAB, space refers to the spatial arrangement of written text: Proximal is where the text starts, and distal is where it ends. However, other correlates of space reflect the physical world rather than arbitrary human conventions of organizing written text. For example, things that are close to us move faster on the retina (a moving car) than things that are farther away (a flying airplane) (Gibson, 1950). Furthermore, things that are close to us are faster to affect us and call for a faster response (Previc, 1990). Because these regularities occur in the real world rather than in written text, they should affect people's cognition irrespective of the prevalent direction of writing in their culture. If CLT is correct in that distances are interrelated and are connected to abstraction, then speed (vs. slowness) would be associated not only with spatial proximity, but also with concreteness, social proximity, temporal proximity and higher likelihood/realism. A series of studies by Nussinson and her colleagues (Nussinson et al., 2022) supported this prediction.

In a study that examined the association between speed and psychological distance, Nussinson and colleagues asked participants to intuitively match each of two words, one reflecting psychological distance (e.g., "theirs", "imaginary") and the other reflecting psychological proximity (e.g., "ours", "real"), to two boxes, one labeled "slow music" and the other "fast music". As hypothesized, participants matched words reflecting psychological distance to the "slow music" box and words reflecting psychological proximity to the "fast music" box with probability that exceeded chance. This held true for each of the four dimensions of psychological distance.

It seems, therefore, that psychological distance versus proximity is indeed experienced as compatible with slowness versus speed, respectively.

A second study by Nussinson and her colleagues examined whether a stimulus psychological distance would affect its perceived speed. Participants read a short description of a location - the city of Jerusalem - described in one condition as located very close to the participants and in the other condition as located very far away from them. (This was made possible because for Israelis who live in Tel Aviv metro area Jerusalem is close if you consider the distance in kilometers, but far away if you consider traffic.) After being induced to think of Jerusalem as being close or far away, participants were asked to imagine the pace of life there, and rate it on a scale from very fast to very slow. As hypothesized, participants who imagined Jerusalem as a spatially distant place portrayed it in their mind's eye as characterized by a slower pace of life compared with participants who imagined it as being spatially close.

Nussinson and colleagues also examined the reverse effect, namely, that of a stimulus' speed on its psychological distance. Participants pictured each of a series of common behaviors (e.g., opening a door) happening either slowly or quickly. Their task was to indicate how imaginary versus real (i.e., how distant on the dimension of hypotheticality) they experienced the behavior to be. As expected, fast behaviors were perceived as being more real than slow behaviors.

Nussinson et al.'s (2022) studies that examined the effect of speed on construal level are also interesting. Participants imagined behaviors occurring slowly or rapidly and then chose between an abstract, high-level description of the behavior focusing on why it is performed and a low-level construal focusing on how it is performed. As expected, participants preferred high- level, "why" descriptions when they imagined the action being performed slowly more than when they imagined the same action being performed quickly.

Would the aforementioned effects connecting speed to distance and construal level depend on horizontal position (left versus right) and writing direction, as SAB would predict? A re-analysis of the data from Nussinson et al.'s paper (2022) showed exactly this. The associations between fast/psychologically close/concrete and between slow/psychologically distant/abstract were stronger in versions of the studies where the concept that represented 'fast' (e.g., faster-pace audios) were presented to the left of the concept that represented 'slow' (e.g., slower-pace audios). Importantly, and also consistent with SAB, this was true in studies con-

ducted in English but was reversed in studies conducted in Hebrew (in which the writing direction is right-to-left). The latter finding, of course, confirms a prediction that is unique to SAB.

It is safe to assume that being fast is more agentic than being slow. That "fast" is close and is located where writing begins is thus consistent with both SAB and CLT. However, the explanation of the association differs between the two frameworks: in CLT, proximity is related to higher speed because spatially proximal objects move faster on the retina. In SAB, on the other hand, action unfolds from where the writing starts, and therefore the left (in right-ward writing cultures) is associated with more agency and speed. It would be interesting to pit these predictions one against the other. For example, it is possible to create stimuli in which the end of a written line appears closer than its origin (see Figure 1B for an example). Who is faster, Ron or Tom? Both frameworks predict that Ron is faster in panel A, but their predictions diverge for panel B, where SAB continues to predict that Ron would be judged as being faster, whereas CLT predicts that Tom would be faster.

Figure 1.A: Left is closer

Figure 1.B: Right is closer

Vertical Position

In our visual field, the lower pane is typically both closer to us and more detailed than the upper pane. The two panes of Figure 2 demonstrate that regularity.

Figure 2. Two scenes that demonstrate the tendency of the upper pane of a visual field to be both less detailed and more distant in space.

Because of that regularity, CLT predicts and a series of studies by Nussinson and colleagues actually find that vertical position is associated not only with abstraction and spatial distance, but also with social distance, temporal distance and hypotheticality (Nussinson et al., 2019, 2021). For example, one study examined the effect of vertical position on temporal distance. Participants were presented with one-sentence scenarios (e.g., "Ron is considering opening a bank account"), and were asked to estimate when in the future the target would perform the activity. Each participant saw sentences presented either at the top or at the bottom of a vertically oriented screen. As hypothesized, participants provided more distant time-estimates when the scenarios were presented at the top of the screen than at the bottom.

Another study examined the effect of vertical position on construal level. It found that positioning personality questionnaire items at the top versus the bottom of a display affected participants' tendency to construe their selves in terms of abstract traits. Were these effects facilitated by writing conventions? Many documents open up with a title or an opening paragraph that convey an abstract gist of their content and proceed with a more specific and detailed description, which naturally appear below. Thus, there seems to be a convention according to which abstract content appears at an upper pane, whereas more concrete content appears at a lower pane.

Consider the following finding: We presented participants with pairs of action descriptions, one of which was more abstract than the other (referred to why the action is done as opposed to how it is done). Participants were instructed to drag and drop each of the two action descriptions into boxes located at the top and bottom of a vertical card presented on a computer screen. As expected, the proportion of description pairs placed in congruence with the hypothesis (the abstract one placed in the upper box) was significantly higher than chance. Could this finding reflect not only the association in our visual system between up and abstract, but also reflect a convention by which abstract is written in an upper pane?

Could it be that the convention reflects the regularity in the visual field? These are important questions for future studies. We think that conventions of writing are not the whole story, because the association also emerged in another study in which participants were presented with one of the two cartoon-like drawings in Figure 3, and had to choose the way the person at the top or the bottom of a mountain would think of an action.
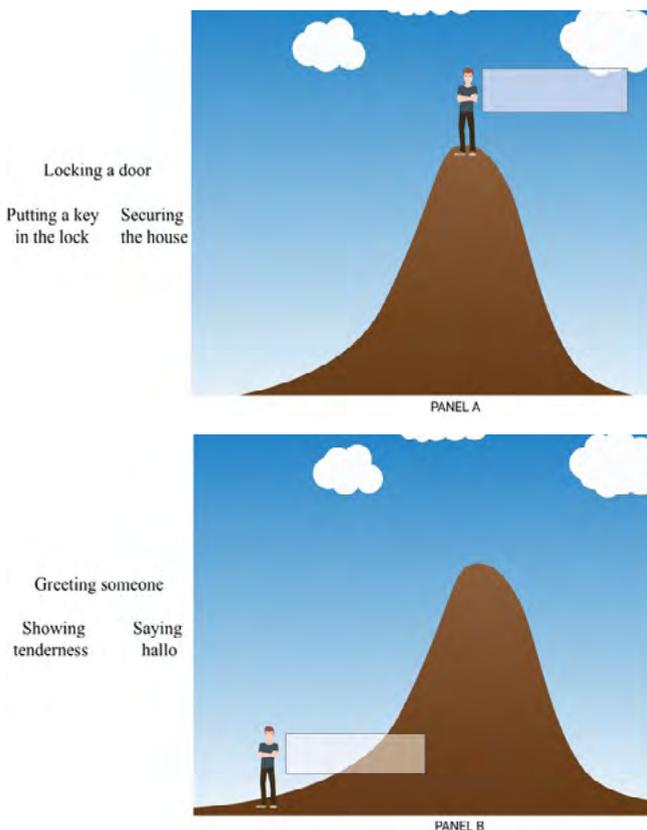

Figure 3. How would a person describe an action? Observers think that it would depend on their vertical location.

As hypothesized, participants were more likely to choose the abstract, why description for the figure at the top of the mountain than at the foot of the mountain. We think that this result would be difficult to attribute solely to writing conventions.

It is interesting to note that writing proceeds downward. In that sense, up is proximal and down is distal, which stands in contrast to the visual regularity (in which, as we noted above, the upper pane typically contains more distal things than the lower one). How this inconsistency would play out in experiments is a question for future studies. For example, some studies by Suitner and Maass have shown that making participants engage in writing exercises that proceed from right to left versus from left to right affect their perception of the agency of left-standing versus right-standing figures. Along similar lines, it would be interesting to examine whether the effects of vertical position on distance and abstraction (Nussinson et al., 2019, 2021) intensify after participants are primed with an "upward-moving" task (e.g., building a tower) and diminish after they are primed with a "downward-moving" task (e.g., dismantling a tower).

Summary and Conclusion

In sum, SAB and CLT frameworks both situate the individual in a grid of distances. Reflecting on the findings of each of these two frameworks from the perspective of the other framework is enriching and elucidating.

For the larger context of embodied cognition, both frameworks suggest that not only the physical world affects our cognition by way of grounding, but also the symbolic, linguistic and metaphoric world affects the way we perceive, understand and reproduce the physical world (Gilead et al., 2021).

References

Amit, E., Rim, S., Halbeisen, G., Priva, U. C., Stephan, E., & Trope, Y. (2019). Distance-dependent memory for pictures and words. Journal of Memory and Language, 105, 119–130.

Fausey, C. M., & Boroditsky, L. (2010). Psychonomic Bulletin and Review, 17, 644- 650.

Formanowicz, M., Roessel, J., Suitner, C., & Maass, A. (2017). Verbs as linguistic markers of agency: The social side of grammar. European

Journal of Social Psychology, 47(5), 566-579.

Gibson, J. J. (1950). The perception of the visual world. Oxford, England: Houghton Mifflin.

Gilead, M., Trope, Y. & Liberman, N., (2021). Grounded separation: Can the sensorimotor be grounded in the symbolic? Behavioral and Brain Sciences. 44, 27-28. DOI:10.1017/S0140525X2000062X

Liberman, N., & Trope, Y. (2008). The psychology of transcending the here and now. Science, 322, 1201–1205.

Liberman, N., & Trope, Y. (2014). Traversing psychological distance. Trends in Cognitive Sciences, 18(7), 364–369.

Liberman, N., Trope Y., & Stephan, E. (2007). Psychological distance. In A. W. Kruglanski & E. T. Higgins (Eds.), Social psychology: handbook of basic principles (pp. 353–383). The Guilford Press.

Maass, A., Ceccarelli, R., & Rudin, S. (1996). Linguistic intergroup bias: Evidence for in-group-protective motivation. Journal of Personality and Social Psychology, 71(3), 512–526. https://doi.org/10.1037/0022-3514.71.3.512

Maass, A., Salvi, D., Arcuri, L., & Semin, G. R. (1989). Language use in intergroup contexts: The linguistic intergroup bias. Journal of Personality and Social Psychology, 57(6), 981–993. https://doi.org/10.1037/0022-3514.57.6.981

Maass, A., Suitner, C., & Nadhmi, F. (2014). What drives the spatial agency bias? An Italian–Malagasy–Arabic comparison study. Journal of Experimental Psychology: General, 143(3), 991–996. https://doi.org/10.1037/a0034989

Nussinson, R., Elias, Y., Mentser, S., Bar-Anan, Y., & Gronau, N. (2019). Bi-directional effects of stimulus vertical position and construal level. Social Psychology, 50(3), 162–173.

Nussinson, R., Elias, Y., Yosef-Nitsan, E., Mentser, S., Zadka, M., Weinstein, Z., & Liberman, N. (2021). Somewhere over the rainbow: An Association between vertical position and psychological distance and construal level. Social Cognition, 39(5), 632-655.

Nussinson, R., Rozenberg, I., Hatzek, A. Mentser, S., Navon, M., Gilead, M, Simchon, A., & Liberman, N. (2022). The poetry of psychological distance: Bi-directional associations between stimulus speed and its psychological distance and construal level. [Manuscript submitted for publication].

Previc, F. H. (1990). Functional specialization in the lower and upper visual fields in humans: Its ecological origins and neurophysiological implications. Behavioral and Brain Sciences, 13, 559–566.

Stephan, E., Liberman, N., & Trope, Y. (2010). Politeness and psychological distance: A construal level perspective. Journal of Personality and Social Psychology, 98(2), 268–280. https://doi.org/10.1037/a0016960

Stephan, E., Liberman, N., & Trope, Y. (2011). The effects of time perspective and level of construal on social distance, Journal of Experimental Social Psychology, 47, 397-402, https://doi.org/10.1016/j.jesp.2010.11.001.

Suitner, C., & Maass, A. (2016). Spatial agency bias: Representing people in space. In J. M. Olson & M. P. Zanna (Eds.), Advances in experimental social psychology (pp. 245–301). Elsevier Academic Press.

Trope, Y., & Liberman, N. (2010). Construal-level theory of psychological distance. Psychological Review, 117, 440–463.

12. Tipping the Balance: Perceptions, Consequences, and Solutions to Economic Inequality

Carmen Cervone[1], Andrea Scatolon[1], Silvia Filippi[1], and Bruno Gabriel Salvador Casara[2]
[1]University of Padova, Italy
[2]New York University Abu Dhabi, United Arab Emirates

Introduction

Economic inequality has quickly become one of the most debated social issues of our times (Piketty, 2020). While interest in this subject has exponentially grown among academics in the past decade, economic inequality has been a topic of discussion since ancient times and also among celebrities and laypeople alike. Think, for example, of Plato (348 B.C.) who believed that "there should exist among the citizens neither extreme poverty nor, again, excessive wealth, for both are productive of great evil", or renowned individuals expressing their concern on how "people can endure such class difference" (Kahlo, 1933), or on how "A nation will not survive morally or economically when so few have so much and so many have so little" (Sanders, 2014).

Surely, economic inequality has been described as the defining social problem of the 21st century (Hauser & Norton, 2017), and for good reason: it is steadily rising (Chancel et al., 2022) and has now reached its historical peak since World War II (OECD, n.d.) - particularly after the Covid-19 pandemic broke out (Berkhout et al., 2021). Mostly, however, the literature argues that economic inequality negatively impacts people

and societies, as (mostly correlational) evidence shows that it is negatively connected to physical and psychological health, as well as societal well-being across the social ladder (Wilkinson & Pickett, 2017; 2020; Oishi & Kesebir, 2015). For instance, on an individual level, higher inequality is connected to decreased life expectancy (De Vogli et al., 2005; Hu et al., 2015), poorer physical health, including life-threatening conditions (Drain et al., 2004; Pickett et al., 2005; Suk et al., 2009), poorer mental health, including lower happiness (Alesina et al., 2004; Oishi et al., 2011; see Filippi et al., 2025 for evidence concerning the link between inequality and well-being at the organizational level), greater status anxiety (Melita et al., 2021; Paskov & Gërxhani, 2013), increased working hours and reduced work-life balance (Filippi et al., 2023), higher depression (Ribeiro et al., 2017), and increased risk behaviors such as drug abuse (Galea et al., 2003). Although Casali et al. (2024) have shown that economic inequality may be linked to certain positive outcomes at the individual level, such as the development of character strengths, these unexpected effects could be explained by increased competitiveness or resilience in the face of adversity, which could potentially be detrimental for people' s health in the long run.

At the societal level, inequality is also connected with bullying (Elgar et al., 2013), dishonesty (e.g., Birkelund & Cherry, 2020; Neville, 2012), corruption (Wei et al., 2022), tax evasion (Bloomquist, 2003; Argentiero et al., 2021), criminality (Kim et al., 2022) and incarceration rates (Phelps & Pager, 2016), lower solidarity and prosociality (Côté et al., 2015; Kirkland et al., 2020; Paskov & Dewilde, 2012; Sandel, 2020), homophobia (Andersen & Fetner, 2008), prejudice (Jetten et al., 2015), lower political engagement (Solt, 2008), and lower voting in elections (Solt, 2010; Wong & Wong, 2022). Some evidence also suggests that inequality has an impact on economic growth and recessions (Cingano, 2014; Piketty, 2014), though some indicates that this is only true for low income countries (Shen & Zhao, 2022), or instead that the relation is positive or inconclusive (Mdingi & Ho, 2021).

So, why is it the case that economic inequality is associated with such a high number of (rather threatening) issues? Two primary paths that link inequality to well-being, mostly at the individual level, have been identified. First, economic inequality is linked to mistrust (Oishi et al., 2011; Uslaner & Brown, 2005), which deters well-being for example by reducing life-satisfaction in the lower classes (Oishi et al., 2011). Second, economic inequality is associated to envy (Cheung & Lucas, 2016), and in

turn, to lower happiness; furthermore, it is associated to value attributed to achievement (Du et al., 2022) and it increases the perception that society is competitive and individualistic (Sánchez-Rodríguez et al., 2019), which instead may contribute to status anxiety. Interestingly, evidence suggests that the first path is more crucial in wealthier societies, whereas the second path in poorer societies. Additionally, another process that has been recently investigated by the literature is conspiratorial thinking, which is made more likely by exposure to high economic inequality (Salvador Casara, Suitner et al., 2022). However, conspiratorial thinking elicited by inequality may ultimately also have a positive effect on society: conspiratorial thinking was shown to be linked with greater support for wealth redistribution through progressive taxation (Salvador Casara et al., 2023).

Finally, the literature argues that in more unequal societies people are more likely to differentiate individuals into solid social classes (Kraus et al., 2017; Browman et al., 2021), making inequality itself more salient. When inequality is more salient and visible, in turn, its negative effects are worse, plausibly because these processes are emphasized: for example, air rage incidents are more common in flights that include first class, especially with front-boarding (DeCelles & Norton, 2016).

While many scholars investigating economic inequality have focused on outcomes on individual and societal wellbeing, current directions in the study of this phenomenon are also examining people's attitudes and preferences towards, and beliefs about, economic inequality.

Perceptions and Attitudes Towards Economic Inequality

Generally, people dislike inequalities. Since childhood, for example, people are inequality averse even at their own cost and/or to the benefit of others (McAuliffe et al., 2017). When it comes to economic inequalities, people believe that current inequalities are too large (OECD, 2021). Furthermore, they desire wealth and income distributions that are more equal than current ones (Norton & Ariely, 2011; OECD, 2021), albeit not perfectly equal (Norton & Ariely, 2011; see Cervone et al., 2021, for similar results on wage distributions). Nevertheless, inequality keeps increasing steadily. This apparent contradiction, of inequality rising even though people are generally against it, was dubbed by researchers the "inequality paradox" (Piff et al., 2018).

Among the first explanations for such a paradox lies the idea that people may accept current inequalities because they are not aware of how large they are. Consistent with this hypothesis, the more inequality people perceive there to be, the more they support governmental redistribution of wealth (Choi, 2019; Gimpelson & Treisman, 2018). Therefore, several studies have investigated whether and how much people misperceive inequalities. Interestingly, people have the notion that economic inequality has been growing in the past years, both in general (OECD, 2021) and through the appraisal of specific timestamps. In two studies, we asked participants to estimate current wealth distribution between the top and bottom wealth quintiles of the population, as well as the wealth distribution ten (Study 1) or five (Study 2) years prior and in the future (Scatolon et al., 2022). Results consistently showed that participants had the accurate perception that inequality has been increasing. However, estimating the actual magnitude of inequality is a different story. Generally, when asked to guess wealth inequality, both adults (Franks & Scherr, 2019) and adolescents (Arsenio & Willems, 2017) underestimate its actual extent. Similar results emerged on pay-gap inequality (Kiatpongsan & Norton, 2014). When it comes to income inequality, instead, evidence shows that individuals overestimate (Chambers et al., 2014) or even correctly estimate (OECD, 2021) country-level income disparities; this may be due to the fact that income inequality is usually smaller than wealth inequality (Piketty & Saez, 2014), and it may also be easier for laypeople to grasp since wealth also includes different, intangible possessions such as assets, stocks, and so on. So, even though results on the direction of the trend are mixed, evidence coming from cross-cultural comparisons (Gimpelson & Treisman. 2018) and methodological reviews (Marandola & Xu, 2021) indicate that overall individuals misperceive economic inequality across different domains.

Nevertheless, correcting misperceptions seems to have only a small overall effect on concern for inequality (see OECD, 2021). Thus, more recent approaches have moved from the misperception approach onto possible psychological strategies that people may employ to legitimize and justify inequality. As previously mentioned, individuals not only reject extreme inequality, but also perfect equality, possibly due to the fact that the latter may be perceived as not rewarding merit and individual differences (see for example the tournament model of wage distributions; Lazear & Rosen, 1981). In this sense, perfect equality may be considered unfair, and indeed some authors have argued that what individuals actu-

ally care about, rather than equality, is fairness (Starmans et al., 2017). In particular, legitimizing beliefs (or legitimization strategies) describe why, to individuals, current inequalities are perceived as fair.

The main theory that, for the past three decades, has been describing why people should endorse legitimizing beliefs is system justification (Jost & Banaji, 1994). According to system justification theory, people are motivated to believe that the systems which they depend on (in this case, the economic system) are fair, because perceived unfairness has a strong psychological toll on the individual. Therefore, legitimizing beliefs carry out the palliative function of lowering cognitive dissonance, guilt, and anxiety deriving from perceived unfairness among both high-status and low-status individuals (Goudarzi et al., 2020; Jost & Hunyady, 2003). While system justification theory in itself is currently under debate (Sotola & Credé, 2022), there is strong evidence that legitimizing beliefs increase acceptance of inequality and decrease support for redistribution. Meritocracy beliefs (i.e., the idea that wealth and poverty result from personal deservingness, and that our society offers equality of opportunity), for example, are linked to acceptance of inequality (García-Sánchez et al., 2019; Schneider & Castillo, 2015) and to lower support for redistribution (Rodriguez-Bailon et al., 2017; Colagrossi et al., 2019). Oppositely, external or situational attributions of poverty (i.e., the belief that poverty is not an outcome of individual action, but of external factors such as corruption) are linked to lower perception that inequality is fair (Schneider & Castillo, 2015), higher support for welfare policies (Bullock et al., 2003) and higher opposition to inequality (Piff et al., 2020). Importantly, these attributions can be actively elicited, for example, through identification with lower-status individuals (Piff et al., 2020).

Similar to meritocracy and causal attributions, beliefs about economic mobility (i.e., that individuals can easily move up and down the social ladder) are connected to economic inequality acceptance (for reviews: Davidai & Wienk, 2021; Day & Fiske, 2019) and perception of its fairness (Day & Fiske, 2017, Heiserman et al. 2020). Specifically, expectations of downward mobility are positively connected to support for redistribution, whereas the opposite is true for upward mobility prospects (Mareeva et al., 2022). However, downward mobility beliefs affect support only for abstract (e.g., general government intervention in redistribution; Steele, 2015) but not concrete (e.g., estate tax policies; Alesina et al., 2018) redistributive policies. Finally, zero-sum beliefs (i.e., perceiving that one party's gains are at another one's losses - particularly, when economic

success is involved; Ongis & Davidai, 2021) and stereotyping and dehumanization of low-class individuals were also evidenced as legitimizing strategies: the former are connected to stronger support for economic inequality (Davidai & Ongis, 2019), whereas the latter have been shown to be linked not only to the rise in economic inequality (Connor et al., 2020; Durante et al., 2013; Sainz et al., 2018) but also to decreased support for welfare policies (regarding healthcare, housing, unemployment, etc.; Sainz et al., 2020).

To sum up, individuals are strongly motivated by their personal beliefs to deem economic inequality as fair, which explains why they may not act to reduce it. Furthermore, this inaction is also reinforced by the fact that people are usually not aware of the actual degree of inequality present in society, and especially in societies where inequality is high, they are less concerned by this issue (Sánchez-Rodríguez & Moreno-Bella, 2021). Thus, given these barriers, how can institutions and individuals promote greater equality?

Restoring the Balance: Redistribution and Collective Action as Tools for Equality

While social psychology has mainly focused on addressing how people perceive and react to inequalities, researchers are now starting to investigate possible solutions to counter this pressing social issue. We will describe here two possible ways, one systemic and one individual, which may lead to the reduction of inequalities and are currently being investigated by socio-psychological literature: redistribution of wealth through taxation and collective action.

How Taxes Can Shape a More Equal Society

International organizations (e.g., Oxfam, 2020) and economists (Pressman, 2014; Piketty, 2021; Stiglitz, 2017) stress the role of taxation, especially if progressive, as a key measure to tackle economic inequality. Despite this, evidence suggests that people are wary of paying taxes (Pew Center, 2020; Ipsos, 2015), especially in contexts characterized by high economic inequality (Salvador Casara et al., 2023). Understanding the psychological processes underlying tax aversion is therefore crucial (Kirchler et al., 2008; Hofmann et al., 2008; Ferrari & Randisi, 2011). In general, citizens' willingness to pay taxes depends on a number of vari-

ables, including internal psychological underpinnings. First, loss aversion (Tversky & Kahneman, 1985) leads people to perceive income as gain and taxes as loss. In fact, taxes are typically associated with a fee or a penalty, and people tend to overlook their positive consequences (Baron & McCaffery, 2005). Second, the misperception of the discrepancy between expected and achieved public services lowers taxpayers' motivation to pay taxes (Lamberton et al., 2018). Along this line, mistrust in institutions appears a key factor of tax aversion (Gangl et al., 2016). Tax knowledge is also an important predictor, since, in absence of it, the tax system may be perceived as excessively complex and cognitively demanding, thereby undermining attitudes towards taxes (Saad, 2014; McKerchar, 2010; Partlow, 2013). Lastly, people's scarce support for redistribution strategies, such as taxes, may depend on a number of individual factors, such as right-wing political ideology or strong meritocracy endorsement (Durante & Fiske, 2017; García-Sánchez et al., 2020; for a review see Hoffman et al., 2008).

The literature suggests that the way redistribution strategies are presented to citizens also plays an important role in shaping attitudes towards them (Fonseca & Grimshaw, 2017; Hallsworth et al., 2017). For example, Baron and McCaffery (2005) showed preliminary evidence of a link between message framing of taxation and attitudes regarding taxes, with participants being susceptible to a variety of heuristics and biases, which occasionally led to incoherent and arbitrary evaluations and judgements. Along the same line, Kamleitner et al. (2012) and Bornman and Wessels (2018) showed that language frames influenced tax compliance, with respondents preferring taxes when described as benefits rather than penalties. Additionally, literature demonstrated that social frameworks and normative compliance messages are useful in increasing good views toward taxes (Cullis et al. 2012; Fonseca & Grimshaw, 2017). Hallsworth et al.'s (2017) natural field experiments highlighted that people were more willing to pay taxes on time when they were exposed to messages that stressed the social norm related to taxes (the majority of people pay taxes on time) and the fact that taxes are fundamental in funding public services that benefit everyone. Compliance may therefore be motivated by the perception of tax usefulness, expressed in terms of its aims (e.g., provision of public services; reducing the gap between rich and poor; Alm et al., 1992). Following this reasoning, Filippi, Suitner, and colleagues (2022) focused on a specific type of frame, pertaining to the aims of taxation and how it is related to construal level, namely the psychological

distance with which taxes can be perceived. Across four studies, authors demonstrated that progressive taxation is perceived as more useful when framed generically, which in turn leads to greater endorsement for this redistribution strategy. Relevant for this there is evidence demonstrating that when the way tax systems work are made transparent and easy to understand for people (via a simulation tool), framing effects disappear (Paetzel et al., 2018).

The way both economic inequality and redistributive policies are framed has important consequences on people's demands for redistribution. For example, Chow and Galak (2012) framed economic inequality as two logically equivalent, yet opposite ways - the rich having more than the poor, or the poor having less than the rich. Results highlighted how presenting information on inequality by employing the former frame strategy resulted in reduced opposition to redistributive tax policies among conservative participants. Employing a similar strategy, Bruckmüller et al. (2017) found that participants perceived bigger differences as less legitimate when these differences were framed as the disadvantaged group having less (see also Bruckmüller, this volume, for how framing inequality as advantage or disadvantage affects responses to inequality more generally). Finally, Dietze and Craig (2020) showed how participants are more willing to support redistribution when the policies are framed as poverty-reducing (i.e., disadvantage-reducing frame) rather than aiming at addressing the gap between the rich and other people (i.e., advantage-reducing frame).

In summary, taxation is an effective strategy to address economic inequalities - but people's support for this redistributive policy is dampened by several factors such as loss aversion, mistrust in institutions, lack of knowledge on the topic, or individual factors such as political ideology. Nevertheless, the way economic inequality and redistributive tax policies are framed play a crucial role in influencing people's attitude on the topic - especially since they can potentially bypass such dampening factors.

Taking a Stand Against Inequality

Collective action has been recently defined as "any actions that individuals undertake as psychological group members to improve the position of a relevant [perceivedly] disadvantaged group as a whole, and/or to protect their or that group's values, moral principles, or ideology" (Agostini & van Zomeren, 2021, p. 685). Collective action allows individuals to exercise change on systems and institutions, and has proven to

be a successful strategy in fighting social inequalities: the Black Lives Matter protests against police brutality in the United States, for example, led to institutional changes such as the reduction of lethal use-of-force by police, which may be partially explained through an increased use of body-cameras and community policing (Campbell, 2021).

When it comes to the economic system, there is empirical reason to believe that a strongly unequal context would facilitate collective action in favor of equality. For example, collective action is strongly based on group identity, perception of injustice, and anger (Agostini & Van Zomeren, 2021). First, economic inequality can foster collective action as it makes wealth categories more salient, thus increasing individual motivations to identify with social groups in terms of economic status: this, in turn, should elicit relative deprivation, perceived injustice, and finally the emotional response of anger. Second, perception of high economic inequality should prompt perception of boundaries between low status and high status groups as impermeable, which, according to Social Identity Theory (Tajfel et al., 1979), should motivate individuals to maintain their positive identity through collective action in favor of equality, rather than through individual social mobility, and to coordinate with the ingroup in order to challenge the status quo (Kawakami & Dion, 1995; van Zomeren et al., 2008). It is not surprising that important movements against economic inequality, such as the Yellow Vest and Occupy Wall Street, emerged in contexts of high economic inequality in order to challenge economic élites. Nevertheless, low perceptions of efficacy (see Agostini & van Zomeren, 2021) against the élites, and the aforementioned legitimization strategies that serve a palliative function against perceived injustices (e.g., system justification theory; Jost et al., 2017), may dampen or counter such processes. How to foster, then, collective action, when legitimization strategies are themselves evoked by its same drivers?

One solution to this conundrum may be provided by a variable that collective action research has been exploring now for the past decade, and has only recently been investigated in relation to economic inequality: morality. According to the dual chamber model of collective action (Agostini & van Zomeren, 2021), morality can motivate collective action by strengthening group identity through a shared norm and by eliciting moral outrage, that is anger experienced as a response to a moral violation. Importantly, we argue that morality (compared to the research on fairness and legitimacy outlined in the previous section) provides a more ample and precise perspective for three primary reasons. First, morality

is subjective, and individuals may not consider fairness to be a relevant principle in their judgments. Relatedly, inequality in particular violates not only fairness principles, but other moral norms as well (e.g., harming low-income people). Thus, reactions to economic inequality may be predicted by a host of moral considerations above and beyond fairness, which (as previously stated) may be more relevant than fairness for certain targets. Finally, morality is at the core of multiple models and theorizations, which have been, and can be, applied to the study of economic inequality.

Previous studies on economic inequality mainly tested the effect of morality on support for redistribution: moral outrage for example was negatively associated with system justification and positively with support for redistribution (Waklsak et al., 2007), and morally framing a charity fighting poverty increased donation intentions towards the charity itself (Franks & Scherr, 2019). In our research, instead, we focused on what drives people to perceive that economic inequality is immoral, and whether this in turns leads to collective action. In two correlational studies, we investigated the perception that specific agents are intentionally causing economic inequality as a potential driver of moralization of inequality, and found that the more people endorsed this belief, the more immoral they believed inequality to be, the more outrage they felt, and the more intention they reported of engaging in collective action (Cervone et al., 2023).

Conclusions

In this chapter, we presented economic inequality as a socio-psychological phenomenon, outlined various ways in which people perceive, conceptualize, and are affected by perceived inequality, and described how we can foster strategies to address it. Given the range and complexity of this phenomenon, studying economic inequality is proving to be a great challenge for researchers, and literature up until now is not without limits. First of all, research has mostly focused on attitudes, preferences and behavioral intentions; real behaviors have instead been under-investigated, with relatively few exceptions (e.g., Piff et al., 2010). Since social psychologists have now started to focus more on possible solutions to this pressing global issue, more attention should be paid to actual behaviors, especially since psychological effects may be weaker on behaviors compared to behavioral intentions (see Agostini & van Zomeren, 2021,

for this finding on collective action). Second, there is still a relative lack of international collaborations on this topic, which in this case is particularly detrimental: inequality and redistribution are issues in which a cross-cultural approach is fundamental, because it allows for ecological comparisons of low- and highly unequal societies, investigations of cultural differences in ideology (e.g., meritocracy beliefs, principles of distributional justice, etc.), and so on. For example, the field could benefit from further comparisons of processes in countries with different levels of inequality (e.g., Durante et al., 2013), or from cross-cultural investigations of tax evasion, which varies greatly from country to country. Additionally, very little attention has been directed to the investigation of cross-country, rather than domestic, economic inequality. Cross-country inequality, however, may have specificities that are not shared with domestic inequality, and thus be a relevant and promising avenue for future research (see for example Vezzoli et al., 2023).

Mostly, we believe that the field is currently too fragmented by micro-theories and constructs. Take, for example, the study of the legitimization of economic inequality: this has been conceptualized through meritocracy beliefs, causal attributions, or beliefs in social mobility, but these constructs can be boiled down to the belief that in our society, people have power over their circumstances. In a similar fashion, the field lacks validated and widely-used measures, and presents instead a tendency to "single-use", ad-hoc items to assess common constructs. For example, Blesch and colleagues (2022) found that contradictory findings in the literature of economic inequality may be due to the predominance of the use of GINI as measure of economic inequality, whereas multi-parameters approaches show better performance.

To conclude, economic inequality is a pervasive and harmful societal issue, and generally recognized as negative by the population. However, efforts to promote equality (for example through redistribution of wealth, or collective action) may be dampened by lack of awareness, as well as system-justifying beliefs that drive individuals to accept or support economic inequality. In this context, social psychology is called to address these challenges and identify strategies to overcome them. To do so, we believe that right now the field needs is an integration of existing perspectives, theories, and evidence into a unitary and cohesive framework. Through this systematic and communitary effort, we believe, researchers can contribute to tipping the scale towards a healthier society.

References

Agostini, M., & van Zomeren, M. (2021). Toward a comprehensive and potentially cross-cultural model of why people engage in collective action: A quantitative research synthesis of four motivations and structural constraints. *Psychological Bulletin*, *147*(7), 667.

Alesina, A., Di Tella, R., & MacCulloch, R. (2004). Inequality and happiness: Are Europeans and Americans different? *Journal of Public Economics, 88*(9–10), 2009–2042. https://doi.org/10.1016/j.jpubeco.2003.07.006

Alesina, A., Stantcheva, S., & Teso, E. (2018). Intergenerational mobility and preferences for redistribution. *American Economic Review, 108*(2), 521-54.

Alm, J., McClelland, G. H., & Schulze, W. D. (1992). Why do people pay taxes?. Journal of public Economics, 48(1), 21-38. doi: https://doi.org/10.1016/0047-2727(92)90040-M

Andersen, R., & Fetner, T. (2008). Economic Inequality and Intolerance: Attitudes toward Homosexuality in 35 Democracies. American Journal of Political Science, 52(4), 942–958. https://doi.org/10.1111/j.1540-5907.2008.00352.x

Argentiero, A., Casal, S., Mittone, L., & Morreale, A. (2021). Tax evasion and inequality: some theoretical and empirical insights. *Economics of Governance*, 1-12.

Arsenio, W. F., & Willems, C. (2017). Adolescents' conceptions of national wealth distribution: Connections with perceived societal fairness and academic plans. Developmental Psychology, 53(3), 463.

Baron, J., & McCaffery, E. J. (2005) Starving the beast: The psychology of budget deficits. Unpublished article. http://dx.doi.org/10.2139/ssrn.589283

Berkhout, E., Galasso, N., Rivero Morales, P. A., Taneja, A., & Vazquez Pimental, D. A. (2021). *The Inequality Virus: Bringing together a world torn apart by coronavirus through a fair, just and sustainable economy.* Oxfam. https://doi.org/10.21201/2021.6409

Birkelund, J., & Cherry, T. L. (2020). Institutional inequality and individual preferences for honesty and generosity. *Journal of Economic Behavior & Organization, 170*, 355-361.

Blesch, K., Hauser, O. P., & Jachimowicz, J. M. (2022). Measuring inequality beyond the Gini coefficient may clarify conflicting findings. Nature Human Behaviour, 1-12.

Bloomquist, K. M. (2003, January). Tax evasion, income inequality and opportunity costs of compliance. In *Proceedings. Annual Conference*

on Taxation and Minutes of the Annual Meeting of the National Tax Association (Vol. 96, pp. 91-104). National Tax Association.

Bornman, M., & Wessels, J. (2018). The tax compliance decision of the individual in business in the sharing economy. eJTR, 16, 425. doi:

Browman, A. S., Destin, M., & Miele, D. B. (2021). The perception of economic inequality weakens Americans' beliefs in both upward and downward socioeconomic mobility. Asian Journal of Social Psychology. DOI: https://doi.org/10.1111/ajsp.12481.

Bruckmüller, S., Reese, G., & Martiny, S. E. (2017). Is higher inequality less legitimate? Depends on How You Frame it!. British Journal of Social Psychology, 56(4), 766-781. doi: https://doi.org/10.1111/bjso.12202

Bullock, H., Williams, W. R., & Limbert, W. M. (2003). Predicting support for welfare policies: The impact of attributions and beliefs about inequality. Journal of Poverty, 7(3), 35–56. Scopus. https://doi.org/10.1300/J134v07n03_03

Campbell, T. (2021). Black Lives Matter's effect on police lethal use-of-force. Available at SSRN 3767097.

Casali, N., Filippi, S. & Feraco, T. Does Inequality Shape Human Character? Cross-Cultural Associations between Character Strengths and the Gini Index in 68 Countries. Journal of Happiness Studies, 25, 37 (2024). https://doi.org/10.1007/s10902-024-00751-w

Cervone, C., Scatolon, A., Lenzi, M., & Maass, A. (2024). Setting Limits: Ethical Thresholds to the CEO-Worker Pay Gap. PLOS One. 19(11): e0311593. https://doi.org/10.1371/journal.pone.0311593

Cervone, C., Suitner, C., Carraro, L., Menini, A., & Maass, A. (2024). Unequal by Malice, Protesters by Outrage: Agent Perception Drives Moralization of, and Collective Action Against, Inequality. British Journal of Social Psychology, 63(4), 1879-1898. https://doi.org/10.1111/bjso.12761

Chambers, J. R., Swan, L. K., & Heesacker, M. (2014). Better off than we know: Distorted perceptions of incomes and income inequality in America. Psychological science, 25(2), 613-618.

Chancel, L., Piketty, T., Saez, E., Zucman, G. et al. World Inequality Report 2022, World Inequality Lab wir2022.wid.world

Cheung, F., & Lucas, R. E. (2016). Income inequality is associated with stronger social comparison effects: The effect of relative income on life satisfaction. *Journal of personality and social psychology*, *110*(2), 332.

Choi, G. (2019). Revisiting the redistribution hypothesis with perceived inequality and redistributive preferences. European Journal of Political Economy, 58, 220-244.

Chow, R. M., & Galak, J. (2012). The effect of inequality frames on support for redistributive tax policies. *Psychological science*, *23*(12), 1467-1469.

Cingano, F. (2014). Trends in Income Inequality and its Impact on Economic Growth (OECD Social, Employment and Migration Working Papers No. 163; OECD Social, Employment and Migration Working Papers, Vol. 163). https://doi.org/10.1787/5jxrjncwxv6j-en

Colagrossi, M., Karagiannis, S., & Raab, R. (2019). The median voter takes it all: Preferences for redistribution and income inequality in the EU 28. (JRC Working Papers in Economic and Finance, Vol. 2019/6). Publications Office. https://data.europa.eu/doi/10.2760/797251

Connor, P., Varney, J., Keltner, D., & Chen, S. (2021). Social class competence stereotypes are amplified by socially signaled economic inequality. Personality and Social Psychology Bulletin, 47(1), 89-105.

Côté, S., House, J., & Willer, R. (2015). High economic inequality leads higher-income individuals to be less generous. Proceedings of the National Academy of Sciences, 112(52), 15838–15843. https://doi.org/10.1073/pnas.1511536112

Cullis, J., Jones, P., & Savoia, A. (2012). Social norms and tax compliance: Framing the decision to pay tax. The Journal of Socio-Economics, 41(2), 159-168. doi: https://doi.org/10.1016/j.socec.2011.12.003

Davidai, S., & Ongis, M. (2019). The politics of zero-sum thinking: The relationship between political ideology and the belief that life is a zero-sum game. Science Advances, 5(12), eaay3761.

Davidai, S., & Wienk, M. N. (2021). The psychology of lay beliefs about economic mobility. Social and Personality Psychology Compass, 15(8), e12625.

Day, M. V., & Fiske, S. T. (2017). Movin'on up? How perceptions of social mobility affect our willingness to defend the system. Social Psychological and Personality Science, 8(3), 267-274.

Day, M. V., & Fiske, S. T. (2019). Understanding the nature and consequences of social mobility beliefs. In The social psychology of inequality (pp. 365-380). Springer, Cham.

DeCelles, K. A., & Norton, M. I. (2016). Physical and situational inequality on airplanes predicts air rage. *Proceedings of the National Academy of Sciences*, *113*(20), 5588-5591.

De Vogli, R., Mistry, R., Gnesotto, R., & Cornia, G. A. (2005). Has the relation between income inequality and life expectancy disappeared?

Evidence from Italy and top industrialised countries. Journal of Epidemiology & Community Health, 59(2), 158-162. DOI: https://doi.org/10.1136/jech.2004.020651

Dietze, P., & Craig, M. A. (2021). Framing economic inequality and policy as group disadvantages (versus group advantages) spurs support for action. *Nature human behaviour*, *5*(3), 349-360.

Du, H., Götz, F. M., King, R. B., & Rentfrow, P. J. (2022). The psychological imprint of inequality: Economic inequality shapes achievement and power values in human life. *Journal of Personality*.

Drain, P. K., Smith, J. S., Hughes, J. P., Halperin, D. T., & Holmes, K. K. (2004). Correlates of National HIV Seroprevalence: An Ecologic Analysis of 122 Developing Countries. *JAIDS Journal of Acquired Immune Deficiency Syndromes*, *35*(4), 407–420. https://doi.org/10.1097/00126334-200404010-00011

Durante, F., & Fiske, S. T. (2017). How social-class stereotypes maintain inequality. Current opinion in psychology, 18, 43-48. doi: https://doi.org/10.1016/j.copsyc.2017.07.033

Durante, F., Fiske, S. T., Kervyn, N., Cuddy, A. J., Akande, A., Adetoun, B. E., ... & Storari, C. C. (2013). Nations' income inequality predicts ambivalence in stereotype content: How societies mind the gap. British Journal of Social Psychology, 52(4), 726-746.

Elgar, F. J., Pickett, K. E., Pickett, W., Craig, W., Molcho, M., Hurrelmann, K., & Lenzi, M. (2013). School bullying, homicide and income inequality: A cross-national pooled time series analysis. *International Journal of Public Health, 58*(2), 237–245. https://doi.org/10.1007/s00038-012-0380-y

Ferrari, L., & Randisi, S. (2011). Psicologia fiscale. Illusioni e decisioni dei contribuenti (pp. 1-249). Raffaello Cortina Editore.

Filippi, S., Salvador Casara, B. G., Peters, K., Maass, A., Feraco, T., & Suitner, C. (2025). They don't really care about us: the impact of perceived vertical pay disparity on employee well-being. European Journal of Work and Organizational Psychology, 34(1), 42-57. https://doi.org/10.1080/1359432X.2024.2415127

Filippi, S., Salvador Casara, B. G., Pirrone, D., Yerkes, M., & Suitner, C. (2023). Economic inequality increases the number of hours worked and decreases work–life balance perceptions: longitudinal and experimental evidence. *Royal Society Open Science*, 10(10), 230187. https://doi.org/10.1098/rsos.230187

Filippi, S., Cervone, C., Maass, A., Del Ben, A., & Suitner, C. (in press). Loving Taxation, Hating Single Taxes: Disentangling Temporal

Distance and Abstraction in the Communication of Tax Proposals. European Journal of Social Psychology, in pressFonseca, M. A., & Grimshaw, S. B. (2017). Do behavioral nudges in prepopulated tax forms affect compliance? Experimental evidence with real taxpayers. Journal of Public Policy & Marketing, 36(2), 213-226. https://doi.org/10.1509/jppm.15.128

Franks, A. S., & Scherr, K. C. (2019). Economic issues are moral issues: The moral underpinnings of the desire to reduce wealth inequality. Social Psychological and Personality Science, 10(4), 553-562.

Galea, S., Ahern, J., & Vlahov, D. (2003). Contextual determinants of drug use risk behavior: a theoretic framework. *Journal of Urban Health, 80*, iii50-iii58.

Gangl, K., Torgler, B., & Kirchler, E. (2016). Patriotism's impact on cooperation with the state: an experimental study on tax compliance. Political Psychology, 37(6), 867-881. doi: https://doi.org/10.1111/pops.12294

García-Sánchez, E., Van der Toorn, J., Rodríguez-Bailón, R., & Willis, G. B. (2019). The Vicious Cycle of Economic Inequality: The Role of Ideology in Shaping the Relationship Between "What Is" and "What Ought to Be" in 41 Countries. Social Psychological and Personality Science, 10(8), 991–1001. https://doi.org/10.1177/1948550618811500

García-Sánchez, E., Osborne, D., Willis, G. B., & Rodríguez-Bailón, R. (2020). Attitudes towards redistribution and the interplay between perceptions and beliefs about inequality. British Journal of Social Psychology, 59(1), 111-136. doi: https://doi.org/10.1111/bjso.12326

Gimpelson, V., & Treisman, D. (2018). Misperceiving inequality. Economics & Politics, 30(1), 27-54.

Goudarzi, S., Pliskin, R., Jost, J. T., & Knowles, E. D. (2020). Economic system justification predicts muted emotional responses to inequality. Nature communications, 11(1), 1-9.

Hallsworth, M., List, J. A., Metcalfe, R. D., & Vlaev, I. (2017). The behavioralist as tax collector: Using natural field experiments to enhance tax compliance. Journal of public economics, 148, 14-31. doi: https://doi.org/10.1016/j.jpubeco.2017.02.003

Hauser, O. P., & Norton, M. I. (2017). (Mis) perceptions of inequality. Current Opinion in Psychology, 18, 21-25.

Heiserman, N., Simpson, B., & Willer, R. (2020). Judgments of economic fairness are based more on perceived economic mobility than perceived inequality. Socius, 6, 2378023120959547.

Hofmann, E., Hoelzl, E., & Kirchler, E. (2008). Preconditions of voluntary

tax compliance: Knowledge and evaluation of taxation, norms, fairness, and motivation to cooperate. Zeitschrift für Psychologie/ Journal of Psychology, 216(4), 209. doi: 10.1027/0044-3409.216.4.209

Hu, Y., Van Lenthe, F. J., & Mackenbach, J. P. (2015). Income inequality, life expectancy and cause-specific mortality in 43 European countries, 1987–2008: a fixed effects study. European journal of epidemiology, 30(8), 615-625. DOI: https://doi.org/10.1007/s10654-015-0066-x

Ipsos, 2015. Governo Italiano: Sondaggi Politico Elettorali. Retrieved 20th July from: http://www.sondaggipoliticoelettorali.it

Jetten, J., Mols, F., & Postmes, T. (2015). Relative Deprivation and Relative Wealth Enhances Anti-Immigrant Sentiments: The V-Curve Re-Examined. PLOS ONE, 10(10), e0139156. https://doi.org/10.1371/journal.pone.0139156

Johnson, W., & Krueger, R. F. (2006). How money buys happiness: genetic and environmental processes linking finances and life satisfaction. Journal of personality and social psychology, 90(4), 680.

Jost, J. T., & Banaji, M. R. (1994). The role of stereotyping in system-justification and the production of false consciousness. British journal of social psychology, 33(1), 1-27.

Jost, J. T., Becker, J., Osborne, D., & Badaan, V. (2017). Missing in (collective) action: Ideology, system justification, and the motivational antecedents of two types of protest behavior. Current Directions in Psychological Science, 26(2), 99-108.

Jost, J., & Hunyady, O. (2003). The psychology of system justification and the palliative function of ideology. European review of social psychology, 13(1), 111-153.

Kahlo, F. (1933). Letter about Fifth Avenue (New York). Retrieved from https://www.liberationnews.org/five-things-to-know-about-frida-kahlo-the-communist/

Kamleitner, B., Korunka, C., & Kirchler, E. (2012). Tax compliance of small business owners: A review. International Journal of Entrepreneurial Behavior & Research. doi: https://doi.org/10.1108/13552551211227710

Kawakami, K., & Dion, K. L. (1995). Social identity and affect as determinants of collective action: Toward an integration of relative deprivation and social identity theories. Theory & Psychology, 5(4), 551-577.

Kiatpongsan, S., & Norton, M. I. (2014). How much (more) should CEOs make? A universal desire for more equal pay. Perspectives on Psychological Science, 9(6), 587-593.

Kim, B., Seo, C., & Hong, Y.-O. (2022). A Systematic Review and Meta-

analysis of Income Inequality and Crime in Europe: Do Places Matter? *European Journal on Criminal Policy and Research, 28*(4), 573–596. https://doi.org/10.1007/s10610-020-09450-7

Kirchler, E., Hoelzl, E., & Wahl, I. (2008). Enforced versus voluntary tax compliance: The "slippery slope" framework. Journal of Economic psychology, 29(2), 210-225. doi: https://doi.org/10.1016/j.joep.2007.05.004

Kirkland, K., Jetten, J., & Nielsen, M. (2020). The effect of economic inequality on young children's prosocial decision-making. British Journal of Developmental Psychology, 38(4), 512-528. DOI: https://doi.org/10.1111/bjdp.12334

Kraus, M. W., Park, J. W., & Tan, J. J. (2017). Signs of social class: The experience of economic inequality in everyday life. Perspectives on Psychological Science, 12(3), 422-435. DOI: https://doi.org/10.1177/17456916166731

Lamberton, C., De Neve, J. E., & Norton, M. I. (2018). The power of voice in stimulating morality: Eliciting taxpayer preferences increases tax compliance. Journal of Consumer Psychology, 28(2), 310-328. doi: https://doi.org/10.1002/jcpy.1022

Lazear, E. P., & Rosen, S. (1981). Rank-order tournaments as optimum labor contracts.Journal of Political Economy,89(5), 841-864

Marandola, G. & Xu, Y. (2021). (Mis-)perception of Inequality: Measures, determinants and consequences. Publications Office of the European Union, Luxembourg. doi:10.2760/444832.

Mareeva, S., Slobodenyuk, E., & Anikin, V. (2022). Support for reducing inequality in the new Russia: Does social mobility matter? *Intersections. East European Journal of Society and Politics*, *8*(2), Article 2. https://doi.org/10.17356/ieejsp.v8i2.782

McAuliffe, K., Blake, P. R., Steinbeis, N., & Warneken, F. (2017). The developmental foundations of human fairness. Nature Human Behaviour, 1(2), 1-9.

McKerchar, M. (2010). Design & Conduct of Research in Tax, Law & Accounting. Thomson Reuters (Professional) Australia Pty Limited.

Mdingi, K., & Ho, S.-Y. (2021). Literature review on income inequality and economic growth. MethodsX, 8, 101402. https://doi.org/10.1016/j.mex.2021.101402

Melita, D., Willis, G. B., & Rodríguez-Bailón, R. (2021). Economic inequality increases status anxiety through perceived contextual competitiveness. Frontiers in psychology, 12, 637365. doi: https://doi.org/10.3389/fpsyg.2021.637365

Neville, L. (2012). Do economic equality and generalized trust inhibit academic dishonesty? Evidence from state-level search-engine queries. *Psychological Science, 23*(4), 339- 345.

Norton, M. I., & Ariely, D. (2011). Building a better America—One wealth quintile at a time. Perspectives on psychological science, 6(1), 9-12.

OECD (n.d.). Inequality. Retrieved 05th October 2022 from https://www.oecd.org/inequality.htm

OECD (2021), Does Inequality Matter?: How People Perceive Economic Disparities and Social Mobility, OECD Publishing, Paris, https://doi.org/10.1787/3023ed40-en.

Oishi, S., & Kesebir, S. (2015). Income inequality explains why economic growth does not always translate to an increase in happiness. Psychological science, 26(10), 1630-1638.

Oishi, S., Kesebir, S., & Diener, E. (2011). Income Inequality and Happiness. Psychological *Science, 22*(9), 1095–1100. https://doi.org/10.1177/0956797611417262

Ongis, M., & Davidai, S. (2021). Personal relative deprivation and the belief that economic success is zero-sum. Journal of Experimental Psychology: General.

Oxfam. Public good or private wealth. In oxfam.org [Internet]. 2020 [cited 2020 October 27]. Available from: https://www.dw.com/en/oxfam-releases-global-inequality-report-amid-ongoing-controversy/a-47142069.

Paetzel, F., Lorenz, J., & Tepe, M. (2018). Transparency diminishes framing-effects in voting on redistribution: Some experimental evidence. European Journal of Political Economy, 55, 169-184. https://doi.org/10.1016/j.ejpoleco.2017.12.002

Partlow, J. (2013). The Necessity of Complexity in the Tax System. Wyoming Law Review, 13(1), 303-334.

Paskov, M., & Dewilde, C. (2012). Income inequality and solidarity in Europe. Research in Social Stratification and Mobility, 30(4), 415–432. https://doi.org/10.1016/j.rssm.2012.06.002

Paskov, M., Gërxhani, K., & van de Werfhorst, H. G. (2013). Income inequality and status anxiety. Growing Inequality Impacts, 90(90), 1-46.

Phelps, M. S., & Pager, D. (2016). Inequality and punishment: A turning point for mass incarceration?. *The ANNALS of the American Academy of Political and Social Science, 663*(1), 185-203.

Pickett, K. E., Kelly, S., Brunner, E., Lobstein, T., & Wilkinson, R. G. (2005). Wider income gaps, wider waistbands? An ecological study of obesity

and income inequality. *Journal of Epidemiology & Community Health, 59*(8), 670–674. https://doi.org/10.1136/jech.2004.028795

Piff, P. K., Kraus, M. W., Côté, S., Cheng, B. H., & Keltner, D. (2010). Having less, giving more: the influence of social class on prosocial behavior. *Journal of personality and social psychology, 99*(5), 771.

Piff, P. K., Kraus, M. W., & Keltner, D. (2018). Unpacking the inequality paradox: The psychological roots of inequality and social class. In Advances in experimental social psychology (Vol. 57, pp. 53-124). Academic Press.

Piff, P. K., Wiwad, D., Robinson, A. R., Aknin, L. B., Mercier, B., & Shariff, A. (2020). Shifting attributions for poverty motivates opposition to inequality and enhances egalitarianism. Nature Human Behaviour, 4(5), 496-505.

Piketty, T. (2014). Capital in the twenty-first century. In *Capital in the twenty-first century*. Harvard University Press.

Piketty, T. (2021). Capital and ideology. In Capital and Ideology. Harvard University Press.

Piketty, T., & Saez, E. (2014). Inequality in the long run. Science, 344(6186), 838-843.

Plato (348 BC). Laws. Wildside Press LLC, 2018.

Pressman, S. (2014). A Tax Reform That Falls Flat. Challenge, 57(4), 82-102. DOI: 10.2753/0577-5132570406

Ribeiro, W. S., Bauer, A., Andrade, M. C. R., York-Smith, M., Pan, P. M., Pingani, L., … & Evans-Lacko, S. (2017). Income inequality and mental illness-related morbidity and resilience: a systematic review and meta-analysis. The Lancet Psychiatry, 4(7), 554-562.

Rodriguez-Bailon, R., Bratanova, B., Willis, G. B., Lopez-Rodriguez, L., Sturrock, A., & Loughnan, S. (2017). Social Class and Ideologies of Inequality: How They Uphold Unequal Societies: Social Class and Ideologies of Inequality. Journal of Social Issues, 73(1), 99–116. https://doi.org/10.1111/josi.12206

Saad, N. (2014). Tax knowledge, tax complexity and tax compliance: Taxpayers' view. Procedia-Social and Behavioral Sciences, 109, 1069-1075. doi: https://doi.org/10.1016/j.sbspro.2013.12.590

Sainz, M., Loughnan, S., Martínez Gutiérrez, R., Moya Morales, M. C., & Rodríguez Bailón, R. M. (2020). Dehumanization of socioeconomically disadvantaged groups decreases support for welfare policies via perceived wastefulness.

Sainz, M., Martínez, R., Moya, M., & Rodríguez-Bailón, R. (2019). Animalizing the disadvantaged, mechanizing the wealthy: The

convergence of socio-economic status and attribution of humanity. International Journal of Psychology, 54(4), 423-430.

Salvador Casara, B. G., Filippi, S., Suitner, C., Dollani, E., & Maass, A. (2023). Tax the élites! The role of economic inequality and conspiracy beliefs on attitudes towards taxes and redistribution intentions. *British Journal of Social Psychology*. doi: https://doi.org/10.1111/bjso.12555

Salvador Casara, B. G., Suitner, C., & Jetten, J. (2022). The impact of economic inequality on conspiracy beliefs. Journal of Experimental Social Psychology, 98, 104245. doi: https://doi.org/10.1016/j.jesp.2021.104245

Sánchez-Rodríguez, Á., & Moreno-Bella, E. (2022). Are you interested in economic inequality? Depends on where you live. *Asian Journal of Social Psychology, 25*(1), 7-19. https://doi.org/10.1111/ajsp.12458

Sánchez-Rodríguez, Á., Willis, G. B., Jetten, J., & Rodríguez-Bailón, R. (2019). Economic inequality enhances inferences that the normative climate is individualistic and competitive. European Journal of Social Psychology, 49(6), 1114-1127. DOI: https://doi.org/10.1002/ejsp.2557

Sandel, M. J. (2020). The tyranny of merit: What's become of the common good?. Penguin UK.

Sanders, B., [@SenSanders], (2014 January 24), *A nation will not survive morally or economically when so few have so much, while so many have so little. #EnoughIsEnough* [Tweet].Twitter. https://twitter.com/sensanders/status/426740006905200640?lang=en

Scatolon, A., Galdi, S., Cervone, C., Mannetti, L., & Maass, A. (2024, February 26). Absolutely Inaccurate, Relatively Precise: Misperceiving Wealth Inequality Across Time. https://doi.org/10.31234/osf.io/fb6ad (è un preprint).

Schneider, S. M., & Castillo, J. C. (2015). Poverty Attributions and the Perceived Justice of Income Inequality: A Comparison of East and West Germany. Social Psychology Quarterly, 78(3), 263–282. https://doi.org/10.1177/0190272515589298

Shen, C., & Zhao, X. (2022). How does income inequality affects economic growth at different income levels? Economic Research-Ekonomska Istraživanja, 1–21. https://doi.org/10.1080/1331677X.2022.2080742

Solt, F. (2008). Economic Inequality and Democratic Political Engagement: ECONOMIC INEQUALITY AND POLITICAL ENGAGEMENT. American Journal of Political Science, 52(1), 48–60. https://doi.org/10.1111/j.1540-5907.2007.00298.x

Solt, F. (2010). Does Economic Inequality Depress Electoral Participation? Testing the Schattschneider Hypothesis. Political Behavior, 32(2), 285–301. https://doi.org/10.1007/s11109-010-9106-0

Sotola, L. K., & Credé, M. (2022). On the predicted replicability of two decades of experimental research on system justification: AZ-curve analysis. European Journal of Social Psychology.

Starmans, C., Sheskin, M., & Bloom, P. (2017). Why people prefer unequal societies. Nature Human Behaviour, 1(4), 1-7.

Steele, L. G. (2015). Income inequality, equal opportunity, and attitudes about redistribution. Social Science Quarterly, 96(2), 444-464.

Stiglitz, J. E. (2017). The overselling of globalization. Business Economics, 52(3), 129-137. DOI: 10.1057/s11369-017-0047-z

Suk, J. E., Manissero, D., Büscher, G., & Semenza, J. C. (2009). Wealth Inequality and Tuberculosis Elimination in Europe. *Emerging Infectious Diseases, 15*(11), 1812– 1814. https://doi.org/10.3201/eid1511.090916

Tajfel, H., Turner, J. C., Austin, W. G., & Worchel, S. (1979). An integrative theory of intergroup conflict. Organizational identity: A reader, 56(65), 9780203505984-16.

Tversky, A., & Kahneman, D. (1985). The framing of decisions and the psychology of choice. In Behavioral decision making (pp. 25-41). Springer, Boston, MA.

Uslaner, E. M., & Brown, M. (2005). Inequality, trust, and civic engagement. *American politics research*, *33*(6), 868-894.

Van Zomeren, M., Postmes, T., & Spears, R. (2008). Toward an integrative social identity model of collective action: a quantitative research synthesis of three socio-psychological perspectives. Psychological bulletin, 134(4), 504.

Veenhoven, R., & Vergunst, F. (2014). The Easterlin illusion: economic growth does go with greater happiness. International Journal of Happiness and Development, 1(4), 311-343.

Vezzoli, M., Valtorta, R.R., Gáspár, A., Cervone, C., Durante, F., Maass, A. & Suitner, C. (2024). Why are some countries rich and others poor? Development and validation of the Attributions for Cross-Country Inequality Scale (ACIS). PLOS One, 19(2): e0298222. https://doi.org/10.1371/journal.pone.0298222.

Wakslak, C. J., Jost, J. T., Tyler, T. R., & Chen, E. S. (2007). Moral outrage mediates the dampening effect of system justification on support for redistributive social policies. Psychological science, 18(3), 267-274.

Wei, C., Dang, J., Liu, L., Li, C., Tan, X., & Gu, Z. (2022). Economic inequality breeds corrupt behaviour. *British Journal of Social Psychology*, bjso.12610. https://doi.org/10.1111/bjso.12610

Wilkinson, R. G., & Pickett, K. (2009). *The spirit level: Why more equal societies almost*

always do better (Vol. 6). London: Allen Lane.

Wilkinson, R. G., & Pickett, K. E. (2017). The enemy between us: The psychological and social costs of inequality. European Journal of Social Psychology, 47(1), 11-24.

Wilkinson, R., & Pickett, K. (2020). *The inner level: How more equal societies reduce stress, restore sanity and improve everyone's well-being.* Penguin.

Wong, M. Y. H., & Wong, S. H.-W. (2022). Income Inequality and Political Participation: A District-Level Analysis of Hong Kong Elections. Social Indicators Research, 162(3), 959–977. https://doi.org/10.1007/s11205-021-02863-9

13. Onward and Upward: Generalization Biases in Inferencing

David L. Hamilton
University of California, Santa Barbara, USA

The two key words in the title of this chapter could characterize forces that propel our everyday lives in important ways. Our lives are guided by well-learned routines as we move through the familiar contexts we encounter, as we anticipate and predict what lies ahead on our path, as we move onward. But our lives involve more than simply experiencing established routines. We encounter new challenges, strive for new goals, and seek to achieve a higher level. We hope to move upward. In this chapter I suggest that similar forces may be important in understanding the inference process. The inferences we make move us forward beyond our current knowledge, and in many contexts they also move us upward to a higher plane, a higher level of understanding.

Within and Between Levels of Representation

We spend a considerable amount of time observing the behaviors of others. When I learn that Kevin got the highest grade in his chemistry course, or when I view Serena making rude remarks to another person, I represent that information in memory. But that is just the beginning of our attempts to understand what we observe. How do we understand Kevin's achievement and Serena's comments, and what do they imply? This is the beginning of the inference process. Inferences go beyond the information at hand, elaborate the initial cognitive representation and provide a basis for anticipating similar behavior in the future. Thus, I

predict that Kevin would also perform well in his calculus class and that Serena will be rude to others as well. These are generalizations anticipating similar performance in similar contexts. They are what we might call horizontal inferences about behaviors at the same level. They are, then, instances of moving onward, going beyond the information given (Bruner, 1957), exploring new horizons.

Although these same-level inferences are useful, the extent of generalization, and hence the breadth of new understanding, is limited. Therefore, we infer properties beyond this behavioral level and infer attributes that are implied by those behaviors. The actor then is seen as possessing more general qualities, moving, in Jones and Davis's (1965) terms, "from acts to dispositions." I not only expect Kevin to do well in other courses but also infer that he is intelligent. Similarly, Serena's rude behaviors lead me to infer that she is a rude person. In doing so my inference is not only characterizing the behavior but also the actor. In contrast to the horizontal inference described above, we can refer to this as a vertical inference. Specifically, it is an inference moving upward to a higher level of abstraction that broadens the domain of generalization.

This transition from the specific (behavior level) to the general (disposition level) is important in several respects. (a) The information perceived gets represented and stored at different levels of representation. Behavior representations are concrete, incorporating features of the act and the context in which it occurred. Abstract representations are more general and are less constrained to specific settings and contexts. Therefore, abstract representations have a broader domain of generalization. Seeing Kevin as an intelligent person, I now expect him to perform well not only in his academic coursework but also in other ways (finding solutions to challenging problems, seeing new insights that are not obvious to many others). (b) Because traits apply to a broader range of trait-relevant behaviors that may occur in a broad range of circumstances, the more abstract level affords a greater range of predictions about the person's behavior. This is useful for guiding one's own behavior in interacting with the person. (c) These traits, inferred from an actor's behaviors, may also be the basis for perceiving similarities and differences between this person and others, for example, mentally grouping Kevin with others who show impressive abilities and Serena with others who seem prone to manifesting socially undesirable behaviors. Once people are grouped into categories on the basis of traits and other generalized qualities, the

foundation has been laid for the formation of stereotypes, which have pervasive implications and consequences.

The transition from the specific to the general in the inference process is moving in an upward direction. However, it is also possible to move between levels in the opposite direction, in a downward direction, inferring behaviors from traits. If I know that Sal is temperamental, I might infer that even a minor affront or comment may "set him off." Maass and colleagues (Maass, Colombo, Colombo, and Sherman, 2001) have referred to these as inductive and deductive inferences, respectively. Clearly people can make both kinds of inferences and both are useful, for different purposes. However, they do not occur equally often or equally easily. Maass et al. (2001) gave participants a series of traits and behaviors descriptive of a person named Marco. Some of the traits were implied by Marco's behaviors, others were not. Participants were later given a recognition test assessing their memory for the traits and behaviors that had been presented. Results showed that people often incorrectly "recognized" traits that had not been presented but were implied by presented behaviors, that is, they frequently moved upward and made inductive inferences. In contrast, deductive inferences (moving downward from traits to behaviors) were infrequent. Moreover, response times for "recognizing" unseen-but-implied traits were as fast as accurate recognition of actually-presented traits, whereas response times for inferred behaviors were quite slow. Maass et al. (2001) identified this as the induction-deduction asymmetry, which has been replicated in subsequent studies (Maass et al., 2001; Maass, Cadinu, Boni, & Borini, 2005; Maass, Cadinu, Taroni, & Masserini, 2006).

Why would this asymmetry occur? In one sense it seems counterintuitive. Knowing that Kevin performed well in one course provides a very narrow basis for assuming that he has the more general trait of intelligence. In contrast, knowing that he is intelligent implies that he would (or at least could) enact a variety of intelligent behaviors in other contexts as well, affording a basis for broader predictions. Still, the Maass at el. (2001) asymmetry exists, and several considerations support it. First, the response time findings point to differences in when these inferences are made. Specifically, inductive inferences seem to occur spontaneously during the encoding of behavioral information, whereas deductive inferences may occur only when activated from memory. Second, traits may be commonly used as organizing units in memory, with corresponding behaviors attached to them (Carlston & Skowronski, 1986;

Hamilton, Katz, & Leirer, 1980; Hamilton, Driscoll, & Worth, 1989; Srull & Wyer, 1989). Entering the network through trait concepts would give the abstract representation priority. Third, traits, being viewed as enduring qualities, may dominate because they are seen as lending stability and coherence to the impressions we form. All of these possibilities would contribute to the inductive-deductive asymmetry.

Upward Inferences: Easy and Pervasive Generalization

We refer to inductive inferences as upward inferences because they move to a higher plane, affording a broader basis for prediction of a person's behavior. Correspondent Inference Theory (Jones & Davis, 1965) focused on the case in which a behavior would induce an inference of an attribute corresponding to the manifest properties of the behavior (inferring from an aggressive behavior that the actor is an aggressive person). According to the theory, such inferences are most likely to occur when the behavior (a) is not constrained by social norms, (b) is undesirable, and (c) the actor had free choice. This analysis suggests that the observer would need to consider these conditions prior to making a correspondent inference. That seems like a lot of cognitive work before inferring that a person's behavior reflects some underlying disposition.

In contrast, research evidence documents that upward (inductive) inferences are made with considerable ease and with few constraints. Maass et al. (2001; Maass et al., 2006) showed that inductive (correspondent) inferences are made quickly, implying they can occur without much thought. In addition, a substantial literature on spontaneous trait inferences (STIs) demonstrates that these inferences occur spontaneously, unintentionally, and without conscious awareness of making them (Carlston & Skowronski, 2005; Hamilton & Thurston, 2023; Todorov & Uleman, 2002, 2003, 2004; Uleman, 1989, 1999; Uleman, Hon, Roman, & Moskowitz, 1996; Uleman, Newman, & Moskowitz, 1996; Uleman, Saribay, & Gonzalez, 2008; Winter & Uleman, 1984). Thus upward inferences are made quickly, especially in comparison to downward inferences. They are, in a word, easy. We can conclude, then, that there is a bias toward upward inferences, toward more abstract levels of comprehension, and toward broader scopes of generalization.

Varieties of Upward Bias

Other research in different domains confirms this upward orientation (see Hamilton & Stroessner, 2021, Chapter 5 for a review). This bias is evidenced in the terms used in identifying individual actions (Vallacher & Wegner, 1987, 1989), the perspective adopted in perceiving behavior (Libby & Eibach, 2011), the linguistic categories used in categorizing behaviors (Semin & Fiedler, 1988, 1992), the characterization of behaviors by ingroup versus outgroup members (Maass, 1999; Maass & Arcuri, 1992), and the psychological distance of a stimulus from the self (Trope & Liberman, 2010). This research has identified variables that influence interpretations that vary along a continuum of very concrete to very abstract interpretations. The level of abstraction makes an important difference in the meaning of the behaviors observed. In the following paragraphs we briefly summarize some examples of this principle.

Linguistic Category Model

When we observe a person's behavior we impose some meaning or interpretation on that behavior. Importantly, the same behavior can be encoded at different levels of abstraction. This is the heart of the Linguistic Category Model (Semin & Fiedler, 1988, 1992). The LCM differentiates several types of verbs that can be used in both describing and in construing behavior: descriptive action verbs (referring to one specific action), interpretive action verbs (referring to a class of behaviors, with positive or negative connotations), and state verbs (referring to the mental state of the actor). These verb types vary along a continuum of concrete (descriptive action) to abstract (state) verb types (Semin & Fiedler, 1992). The highest level of abstraction in LCM is adjectives (referring to properties of the actor but not to specific actions). Thus, for example, Marcus hit Tony, Marcus hurt Tony, Marcus hates Tony, and Marcus is aggressive. Others have argued that nouns are an even higher level of abstraction and convey more about the target person than the other categories (Carnaghi, Maass, Gresta, Bianchi, Cadinu, & Arcuri, 2008; Hamilton, Gibbons, Stroessner, & Sherman, 1992; Maass, Carnaghi, & Rakic, 2015). Describing Marcus as a Jew is a more abstract – broader, more pervasive and inclusive -- characterization than saying he is Jewish. Moreover, a person referred to by a noun category (e.g., athlete), as opposed to by an adjective (athletic) is expected to perform more stereotypic and fewer counterstereotypic behaviors. Thus a linguistic cue (noun vs. adjective

descriptor) can serve to maintain stereotypical expectations in everyday communication (Maass et al., 2015).

The LCM shows that aspects of language use -- different verb forms, adjectives, and nouns, all of which might pertain to the same referent (behavior, person) -- can guide and influence one's understanding of the target of the description. Variation of these descriptors along the concrete-abstract dimension affects perceptions of the duration of a described behavior, how enduring a quality of a person is, the attributions made and how easy or difficult it is to identify, verify, confirm, and disconfirm these behaviors (Semin & Fiedler, 1992). All of these language properties can influence the extent and breadth of inferences drawn from the information conveyed. The higher the level of abstraction, the more general the characterization and hence the more stable, enduring, and unchangeable the quality seems to be.

Linguistic Intergroup Bias

These linguistic effects are also manifested in intergroup contexts. People typically feel more favorable toward groups to which they belong (ingroups) than toward groups they are not a part of (outgroups), known as ingroup bias. Maass and her colleagues (Maass, Salvi, Arcuri, & Semin, 1989) showed that this bias can be enhanced and perpetuated in language use, known as the Linguistic Intergroup Bias (LIB). They presented participants drawings of ingroup and outgroup members (rival teams) enacting either desirable or undesirable behaviors. For each one, participants had to select one of four descriptions that best characterized the behavior. The four descriptors were worded according to the four levels of the LCM. When ingroup members performed desirable behaviors the participants chose higher, more abstract descriptors than when outgroup members performed desirable behaviors. For undesirable behaviors the opposite pattern cooccurred; higher generality was selected in characterizing outgroup than ingroup negative behaviors. Thus the overall upward bias is moderated by the ingroup/outgroup status of the actor. Obviously, this bias in characterizing ingroups versus outgroups is important because abstract language implies greater temporal stability and cross-situational consistency and hence more enduring qualities of the target person. Therefore the LIB serves to perpetuate perceptions of intergroup differences and stereotypes.

The LIB has been replicated many times, with many kinds of ingroup-outgroup distinctions, and in many social contexts (Anolli, Zur-

loni, & Riva, 2006; Arcuri, Maass, & Portelli, 1993; Karpinski, & Von Hippel, 1996; Maass, 1999; Maass & Arcuri, 1992, 1996; Maass, Ceccarelli, & Rudin, 1996; Maass, Corvino, & Arcuri, 1994; Maass, Milesi, Zabbini, & Stahlberg, 1995; Shulman, Collins, & Clement, 2011; Tincher, Lebois, & Barsalou, 2016). The effect can sometimes appear more strongly in some groups than others (Schulman et al., 2011; Semin, de Montes, & Valencia, 2003) and the effect can to some extent be diminished (but not eliminated) by a brief mindfulness training (Tincher et al., 2016).

The LIB can be communicated in print media. Content analyses of newspaper articles about immigration showed that comments describing the positive ingroup (U.S.) used more abstract language than did comments about an outgroup (undocumented immigrants) and negative ingroup comments. However, negative comments about the outgroup used more abstract language than did positive comments about the outgroup or negative comments about the ingroup (Dragojevic, Sink, & Mastro, 2017; Maass et al., 1994). Moreover, this pattern of language use resulted in more unfavorable attitudes toward Latinos (Mastro, Tukachinsky, Behm-Morawitz, & Blecha, 2014). Thus, the LIB bias in language use can contribute to the persistence of group stereotypes.

What processes underlie the LIB? Two possibilities have been investigated. A cognitive explanation argues that the LIB derives from differential expectancies. Expectancy congruent behaviors are described in abstract terms because they are seen as reflecting stable qualities of the actor. Expectancy inconsistent behaviors are viewed as atypical, not reflecting enduring qualities, and therefore are interpreted in more concrete terms. Together, these tendencies produce the LIB. A second explanation rests on ingroup protection motives. People are motivated to feel good about the groups they belong to and draw some positive self-regard from their membership groups. One way to achieve that is to differentiate one's own group from an outgroup. The LIB may serve that goal by upwardly interpreting positive ingroup behaviors and negative outgroup behaviors in terms of abstract (stable, enduring) qualities and using concrete encodings of negative ingroup and positive outgroup behaviors. It may be, of course, that both mechanisms can contribute to the LIB, and both explanations have received some empirical support (Karpinski & Von Hippel, 1996; Maass et al., 1995; Maass et al., 1996; Wigboldus, Semin, & Spears, 2000).

In sum, the LIB is a highly robust phenomenon that can influence both the level of abstractness with which perceivers encode and interpret

behaviors as well as the abstractness of language employed in communicating behavior descriptions to others.

Cultural Differences in Manifestations of Generalization Bias

The evidence documenting these biases in inferences is pervasive, sufficient to make one wonder if they are universal. Relevant to that question is research on the role of culture on these effects. Although the amount of research comparing peoples from different cultures has mushroomed in recent decades, research specifically investigating inference processes is not large. Nevertheless, what evidence we have clearly suggests cultural differences in the nature of inferences as they are commonly made. Most of that research has focused on comparing people living in or emanating from East Asian and Western cultural contexts and has explored differences in language use.

Cultural differences in construing social behavior and person perception have been well studied and have identified differences in cognitive functioning. One approach views the distinction as between individualistic (Western) and collectivist (Eastern) cultures. The individualistic perspective sees the individual person as independent from others and emphasizes personal qualities such as traits and other internal attributes that are stable across time and situations. The collectivist view sees people as interdependent, immersed in and defined by their social relationships, emphasizing the importance of social roles and sensitive to the social context (e.g., Markus & Kitayama, 1991; Triandis, 1989).

An alternative approach to construing cultural differences focuses on a difference in cognitive styles between Eastern and Western cultures. Westerners tend to use an analytic cognitive style, which views the focus of attention (person, group) as detached from its context and instead emphasize the attributes of that focal object. Easterners, in contrast, tend to use a holistic cognitive style that focuses on the social context and relationships within that context, understanding behavior and events in terms of those relationships (Masuda & Nisbett, 2001; Nisbett, Peng, Choi, & Norenzayan, 2002).

These cultural differences are manifested in the ease and likelihood of making spontaneous trait inferences (STIs). Several studies have shown that, although both Eastern and Western samples make STIs, they tend to be more frequent and stronger among individualistic European American samples than among collectivist Asians or Asian American samples (Lee,

Shimizu, Masuda, & Uleman, 2017; Na & Kitayama, 2011; Shimizo, Lee, & Uleman, 2017; Zarate, Uleman, & Voils, 2001).

STI is clearly an instance of upward generalization, an inference from the specific behavior to the more abstract dispositional quality of the actor. However, when people observe the behavior of others they witness more than the behavioral act performed by the actor. They also see the social context in which the behavior occurs, and that context has properties and attributes of its own. Some research has argued that perceivers also may make spontaneous situational inferences (SSIs) in processing social information (Ham & Vonk, 2003; Knowles, Morris, Chiu, & Hong, 2001; Newman & Marsden, 2023; Ramos, Garcia-Marques, Hamilton, Ferreira, & Van Acker, 2012). Moreover, the relation between STIs and SSIs is not an either/or distinction, as both kinds of spontaneous inferences can occur simultaneously (Ham & Vonk, 2003; Todd. Molden, Ham, & Vonk, 2011). But are they equally likely to occur in both individualistic and collectivist cultures? In fact, the analytic-holistic distinction in cognitive style noted above suggests that Easterners (adopting the holistic style) are more attuned to those situation properties. Lee et al. (2017) examined that question in a study that measured both STIs and SSIs in both European Canadian and Japanese samples. They found that both groups made both STIs and SSIs, but there were cultural differences in their manifestation. The European Canadians made more STIs than did the Japanese participants but there was no cultural difference in SSIs.

The analytic cognitive style prominent in Western cultures emphasizes the importance of traits (and other internal qualities) in understanding another's behavior, more so than the holistic style that is more typical of Eastern cultures. Maass, Karasawa, Politi, and Suga (2006) studied the implications of this difference in language use in the context of the LCM (Semin & Fiedler, 1988). They had samples of Italian and Japanese participants write descriptions of individual persons they know well and of men and women in general. In each case they were instructed to list 10 aspects that are descriptive of the target. These descriptions were coded according to the categories of the LCM. The analyses showed that Italians used many more adjectives than verbs in their descriptions, whereas Japanese made more frequent use of verbs in their descriptions. Thus, a more abstract level of representation is communicated in the language produced by European than Asian groups.

Spontaneous Generalization Across Group Members

The research we have summarized reveals an upward bias in inference processes. In all of this research the target of inference has been an individual. Yet we routinely observe people in groups and make inferences about those groups. Are there similar biases in people's inference processes about groups?

Maass et al.'s (1989) research demonstrated the role of language in perpetuating the in-group bias, evaluatively favoring ingroups over outgroups. In their studies the stimulus persons were members of well-known ingroups and outgroups, often in direct competition with each other. Other research has shown that the ingroup bias occurs even in judgments of ingroup and outgroup members of minimal groups, when the participants do not even know the specific basis of the distinction between the groups (Brewer, 1979; Tajfel, 1970; Tajfel Billig, Bundy, & Flament, 1971). This raises a broader question about the ingroup bias.

We know that people make inferences from behaviors to traits spontaneously, without conscious intent or awareness (STIs). Is the ingroup bias a result of deliberative thought and comparison between ingroup and outgroup? Or can ingroup favoritism occur spontaneously, simply as people process information about own- and other-group members? In one study (Otten & Moskowitz, 2000) participants were randomly assigned to arbitrary groups, presumably on the basis of different perceptual styles. They then read trait-implying behaviors, both positive and negative in valence, performed by ingroup and outgroup members. Following each sentence a positively or negatively valenced trait word was presented. Sometimes the trait was implied by the behavior, sometimes not, and participants' were to indicate whether or not the trait word had appeared in the sentence. In every case the correct answer was No. However, if an STI had been made as the behavior was encoded, the word would be highly accessible, would create uncertainty as to whether it had actually been seen or was merely assumed, and this uncertainty would slow down response times in answering the question. Analyses of these response times showed that STIs were made (i.e., slower decision times) when sentences described ingroup members performing positively-valenced behaviors. Thus ingroup favoritism was manifested in STIs when behaviors by ingroup members were processed. This result corresponds to Maass et al.'s (1989) finding that more abstract descriptors were chosen when ingroup team members performed desirable behaviors.

Onward and Upward Generalization in Inferences About Group Members

In both of the studies described above participants were responding to information about individual group members and the results showed that their ingroup/outgroup status influenced the abstractness of descriptors chosen by participants to characterize them. But does the effect on spontaneous judgments of individual group members generalize to the group as a whole? We know that observers spontaneously infer a person's attributes from that person's behaviors (STIs). Groups are comprised of individual group members, so the question naturally arises whether the inferences inferred, when made about individual group members, generalize to other members of that group. If so, then members of the group would be perceived as having the same or similar attributes, even though each of those attributes was inferred about only one of the group's members. This process would produce an exaggerated perception of similarity among the group members, more so than warranted by the information acquired about them. This would be important because heightened perceptions of similarity among group members can be an important precursor to the development of group stereotypes.

Of course, there are many types of groups and it may be that generalization is more likely in some groups than in others. The term entitativity refers to the perception that a collection of persons constitutes a single entity, a group (Campbell, 1958; Hamilton, Chen, & Way, 2015). For example, those persons may share some similarities -- inherent (e.g., race, gender, age) or acquired (e.g., uniforms) features -- that promote the appearance of being a group. Similarly, if persons interact a lot, if those interactions appear coordinated as if in pursuit of common goals, or if they share a common outcome (e.g., sports teams), they are more likely to be perceived as a meaningful group. Groups can vary in all of these properties, thus differing in perceived entitativity. Once entitativity is perceived, the group is seen as having unity and coherence, which facilitates drawing inferences about the group and its members. These differences in perceived entitativity might influence the likelihood that STIs about group members would be generalized across group members.

These ideas were tested in a study (Crawford, Sherman, & Hamilton, 2002) in which participants were presented a series of stimuli, each consisting of a face photo with a behavior performed by that person. The stimulus persons were identified as members of two groups, one of which was described as a tight-knit group in which members shared common

goals (high entitativity) whereas members of the other group were more loosely connected (low entitativity). The behaviors by members of one group implied different traits than did the behaviors of the other group's members. Later the same stimulus persons were shown again, each of them paired with a trait word. In some cases the trait was implied by that person's behavior whereas in other cases the trait was implied by the behavior of a different member of the same group. Finally, the faces were shown again and, for each one, participants' task was to recall the trait what had been paired with that face in the previous phase. The question of interest was whether traits inferred about one group member would be spontaneously transferred to other members of the same group. This transfer did occur, but only when the group was high in entitativity (Hamilton, Sherman, & Rodgers, (2004). Thus traits inferred about group members generalized to other members of the same group (but not to a different group). When this happens the members of the same group are perceived as more similar to each other, the group members become more interchangeable, increasing perceptions of within-group homogeneity, which is an important precondition for stereotype formation. Importantly, all of this has happened spontaneously, without intention, and without the perceiver's awareness that it has happened. Again, when a group was high in entitativity, the inference process has moved both onward and upward. It has moved onward as the inferred attributes have generalized across group members (same level), assimilating different group members to each other by their shared attributes, and it has moved upward to a more general level as the inferred attributes create a more abstract representation of the group. This enhanced perception of similarity at an abstract level can lay the groundwork for stereotype formation.

Spontaneous Trait Inferences About Groups (STIGs)

In the studies just described, the stimulus information consisted of behavioral portrayals of individual persons -- visual episodes in cartoon-like form (Maass et al., 1989), faces with behavior descriptions (Crawford et al., 2002) – and inferences from this individual-based information has influenced evaluations and attribute representations of the groups of which those persons were members. But what about the groups themselves? We regularly observe groups acting as a unit. A student group demonstrated in protesting a university policy. A church group volunteered at a soup kitchen. We observe these behaviors being performed by a group

as an entity, not as a collection of individuals. From these observations, do observers spontaneously infer that these groups, as groups, possess corresponding attributes? Hamilton, Chen, Ko, Winczewski, Banerji, and Thurston (2015) investigated this question and found that people do make spontaneous trait inferences about groups (STIGs), which can have important effects on perceptions of those groups.

In a series of studies participants were shown a number of stimuli, each showing a set of four faces (groups) and a sentence describing a group action. After this presentation the sets of four faces were shown again with a trait word. The trait either was implied by the group's behavior (Match trial) or was implied by the behavior of a different group (Mismatch). Participants' task was to indicate whether or not the trait word was in the sentence that described that group. In all cases the answer was No. If, however, participants had made an STI while encoding the behavior they would be more likely to say Yes to the probe question (that is, it would be a false recognition). If false recognitions occur more frequently on Match than on Mismatch trials, it would be evidence that STIGs had been made during encoding.

Results of several experiments documented that STIGs are made while processing group-relevant behavior. The initial study (Hamilton et al., 2015, Experiment 1) compared STIs for individual targets with STIGs for group targets and found that both kinds of spontaneous inference were made, using the same materials and procedures except for the different targets. Another study (Hamilton et al., 2015, Experiment 2) compared evidence of STIGs when participants were or were not under cognitive load. Participants again made STIGs but the two load conditions did not differ, indicating that STIGs, being spontaneous during encoding, do not require a lot of cognitive resources. Results from these studies document that STIGs do happen in processing information about groups. Once again, perceivers are spontaneously moving upward to a more general, abstract level of understanding.

One might wonder, though, how general this effect is. Do people make STIGs for all groups, or just certain groups, or certain kinds of groups? Given the range and diversity of groups, it's not clear how to approach those questions. As a first attempt, the next study (Hamilton et al., 2015, Experiment 3) investigated STIGs in processing information about groups high and low in entitativity. At first glance one might expect more evidence of STIGs for high (close, coherent units) than -for low (not tightly connected) entitativity groups. People process information about high

entitativity groups much as they process information about individual persons (for whom STIs are commonly made), but less so for low entitativity groups that lack unity and coherence (Hamilton & Sherman, 1996). On the other hand, if STIGs are made spontaneously during encoding, they may routinely occur for all groups. In the study the perceived entitativity was manipulated through instructions, and then participants learned about a series of groups by viewing photo + behavior descriptions of the groups. Results showed that STIGs were made for both high and low entitativity groups, but there was no difference in generation of STIGs in the two conditions. This result suggests that STIGs were made spontaneously during the encoding of group information for all groups, regardless of entitativity. Further research is needed to explore this hypothesis with different types of groups and different means of operationalizing entitativity.

Given the spontaneous nature of STIGs, how do they influence subsequent processing and perceptions? Being made during the initial comprehension of behavioral information, they would lay the foundation for forming impressions, in this case impressions of group targets. In their next study Hamilton et al. (2015, Experiment 4) again presented the same information about a series of groups. Then in the second phase participants were shown the group faces again but instead of the recognition task participants rated their impression of each group on a series of traits. Analyses showed that participants rated each group higher on traits implied by that group's behavior (unique to each group) than on other (non-implied) traits. Thus, STIGs – unintended, nonconsciously-produced inferences – can be the basis for group impressions.

In its fundamental elements, stereotypes are impressions of groups. They represent the important attributes believed to characterize various groups. The fact that group impressions can emerge from inferences spontaneously made from behaviors of groups suggests that the initial foundation for development of group stereotypes may occur spontaneously, without intent or conscious awareness – an intriguing possibility in its own right. But stereotypes are more than simple lists of traits associated with a group. One of the important contributions of Allport's (1954) classic analysis was his observation that stereotypes, once formed, are applied to all members of the group. Allport emphasized the routine generalization of the stereotype to newly-encountered group members. As we described above, Crawford et al (2002) documented the generalization of attributes inferred from specific behaviors of group members

to a more abstract representation in which the inferred traits are generalized to other group members. In addition, Allport's observations suggest downward generalization from abstract representation to individual group members. In other words, whereas Crawford et al. (2002) demonstrated what Maass et al. (2001) called inductive inference, Allport's analysis suggests that stereotyping might also occur spontaneously through deductive inference.

Hamilton et al. (2015, Experiment 5) adapted their paradigm to test these ideas. Participants again were shown a series of photos of four-person groups and a sentence describing a behavior performed by each group. Then, as in the earlier studies, the group faces were shown again, each one with a trait word, and participants indicated whether or not that word had been in the sentence describing that group. Again, the word had not been in the sentence, but a Yes answer (a false recognition) would indicate that a spontaneous trait inference had been made. The next phase was new to this study. Participants were shown a photo of a new member of the group, not previously seen, with no information about him (other than his group membership), and they were asked to rate him on the same traits used in the previous experiment. These ratings showed that the new member was rated higher on traits implied by the group's behavior than on other traits (implied by the behavior of other stimulus groups). Thus, STIG-based inferences generalized to a new member of the group about whom no information had been provided.

Implications for Stereotype Formation: A New Process

These experiments investigating STIGs have extended research on spontaneous inferences in several ways. They demonstrate that spontaneous inferences are made as perceivers process information about behaviors enacted by groups, that STIGs occur for different types of groups and even when processing under cognitive load, that STIGs provide the basis for first impressions of groups, and that STIGs generalize to a new, unknown group member.

The evidence that group impressions may come about through STIGs suggests a new cognitive mechanism by which stereotypic concepts may form (Hamilton et al., 2015). The process would begin with observation of group behavior, producing a STIG, thereby inferring a disposition characterizing the group. Thus the group is endowed with qualities imposed by the STIG process. This means that a stereotypic concept is formed with-

out conscious intention or conscious awareness, yet it becomes a part of the perceiver's mental representation of the group. Once established in memory, these initial conceptions can be elaborated, strengthened, and perpetuated by the same cognitive and motivational processes (and biases) that maintain other cognitive representations. Therefore STIGs – spontaneously inferred group attributes – may lay the foundation for a newly formed stereotype.

Historically, most explanations for the development of stereotypes have given important roles to other factors: a history of animosity and conflict between groups, the interdependence between groups and competition for scarce resources, the relative deprivation between one's own and a target group, and comparisons between groups. Note that none of those preconditions were present in these STIG experiments. Rather, group concepts were just born spontaneously and without intention or awareness of their emergence. Yet the early fragments of group impressions were evident in the findings. These ideas are, of course, somewhat speculative and the implications of this research on STIGs for the emergence of stereotypic conceptions needs to be developed and empirically tested. However, they are supported by the evidence available and warrant further exploration.

Concluding Thoughts

One purpose of this book is to bring together several perspectives for honoring the rich and varied career of Anne Maass. This chapter is a contribution to that effort. We have discussed a number of topics that range widely across the domains of person and group perception, highlighting the variety of ways Anne's thinking has enhanced our science. The remarkable fact is that virtually all of the work I have presented mirrors or was (directly for indirectly) stimulated by research done by Anne Maass. Her writings have shaped how we think about these topics and her theoretical ideas and empirical research have inspired research by other investigators. That has certainly been true of Anne's influence on me. I am honored to be a part of this volume devoted to recognizing this remarkable person and her outstanding career.

Note: The author is grateful to Yoshi Kashima for his very useful comments on an earlier version of this chapter.

References

Anolli, L., Zurloni, V., & Riva, G. (2006). Linguistic intergroup bias in political communication. *Journal of General Psychology, 133*, 237-255.

Arcuri, L., Maass, A., & Portelli, G. (1993). Linguistic intergroup bias and implicit attributions. *British Journal of Social Psychology, 32*, 277-285.

Brewer, M.B. (1979). Ingroup bias in the minimal intergroup situation: A cognitive-motivational analysis. *Psychological Bulletin, 86*, 307-324.

Brewer, M.B. (2015). Motivated entitativity: When we'd rather see the forest than the trees. In S.J. Stroessner & J.W. Sherman (Eds.), *Social Perception: From Individuals to Groups* (pp. 161-176). New York: Psychology Press.

Bruner, J.S. (1957). On perceptual readiness. *Psychological Review, 64*, 123-152.

Campbell, D.T. (1958). Common fate, similarity, and other indices of the status of aggregates of persons as social entities. *Behavioral Science, 3*, 14-25.

Carlston, D.E., & Skowronski, J.J. (1986). Trait memory and behavior memory: The effects of alternative pathways on impression judgment response times. *Journal of Personality and Social Psychology, 50*, 5-13.

Carlston, D.E., & Skowronski, J.J. (2005). Linking versus thinking: Evidence for the different associative and attributional biases of spontaneous trait inference and spontaneous trait transference. *Journal of Personality and Social Psychology, 89,* 884-898.

Carnaghi, A., Maass, A., Gresta, S., Bianchi, M., Cadinu, M., & Arcuri, L. (2008). Nomina sunt omina: On the inductive potential of nouns and adjectives in person perception. *Journal of Personality and Social Psychology, 94*, 839-859.

Crawford, M.T., Sherman, S.J., & Hamilton, D.L. (2002). Perceived entitativity, stereotype formation, and the interchangeability of group members. *Journal of Personality and Social Psychology, 83,* 1076-1094.

Dragojevic, M., Sink, A., & Mastro, D. (2017). Evidence of linguistic intergroup bias in U.S. print news coverage of immigration. *Journal of Language and Social Psychology, 36,* 462–472.

Ham, J., & Vonk, R. (2003). Smart and easy: Co-occurring activation of spontaneous trait inferences and spontaneous situational inferences. *Journal of Experimental Social Psychology, 39*, 434-477.

Hamilton, D.L., Chen, J.M., & Way, N. (2011). Dynamic aspects of entitativity: From group perception to social interaction. In R.M. Kramer, G.J. Leonardelli, & R.W. Livingston (Eds.), *Social Cognition,*

Social Identity, and Intergroup Relations: A Festschrift in Honor of Marilynn Brewer (pp. 27-52). New York, NY: Psychology Press.

Hamilton, D.L., Chen, J.M., Ko, D., Winczewski, L., Banerji, I., & Thurston, J. A. (2015). Sowing the seeds of stereotypes: Spontaneous inferences about groups. *Journal of Personality and Social Psychology, 109,* 569-588.

Hamilton, D. L., Driscoll, D. M., & Worth, L. T. (1989). Cognitive organization of impressions: Effects of incongruency in complex representations. *Journal of Personality and Social Psychology, 57,* 925-939.

Hamilton, D. L., Gibbons, P., Stroessner, S. J., & Sherman, J. W. (1992). Stereotypes and language use. In G. R. Semin & K. Fiedler (Eds.), *Language, Interaction, and Social Cognition* (pp. 102-128). London: Sage.

Hamilton, D. L., Katz, L. B., & Leirer, V. O. (1980). Organizational processes in impression formation. In R. Hastie, T. M. Ostrom, E. B. Ebbesen, R. S. Wyer, Jr., D. L. Hamilton, & D. E. Carlston (Eds.), *Person Memory: The Cognitive Basis of Social Perception* (pp. 121-153). Hillsdale, NJ: Lawrence Erlbaum Associates.

Hamilton, D. L., & Sherman, S. J. (1996). Perceiving persons and groups. *Psychological Review, 103,* 336-355.

Hamilton, D.L., Sherman, S.J., & Rodgers, J.S. (2004). Perceiving the groupness of groups: Entitativity, homogeneity, essentialism, and stereotypes. In V. Yzerbyt, C.M. Judd, & O. Corneille (Eds.), *The Psychology of Group Perception: Perceived Variability, Entitativity, and Essentialism* (pp. 39-60). New York, NY: Psychology Press.

Hamilton, D.L., & Stroessner, S.J. (2021). *Social Cognition: Understanding People and Events.* London: Sage.

Hamilton, D.L., & Thurston, J.A. (2023). Perceiving group attributes spontaneously: Broadening the domain. In E. Balcetis & G.B. Moskowitz (Eds.), *The Handbook of Impression Formation: A Social Psychological Approach* (pp. 324-347). New York: Routledge.

Jones, E.E., & Davis, K.E. (1965). From acts to dispositions: The attribution process in person perception. In L. Berkowitz (Ed.), *Advances in Experimental Social Psychology* (Vol. 2, pp. 219-266). New York, NY: Academic Press.

Karpinski, A ., & Von Hippel, W. (1996). The role of linguistic intergroup bias in expectancy maintenance. *Social Cognition, 14,* 141-163.

Knowles, E.D., Morris, M.W., Chiu, C.-Y., & Hong, Y.-Y. (2001). Culture ands the process of person perception: Evidence for automaticity

among East Asians in correcting for situational influences on behavior. *Personality and Social Psychology Bulletin, 27*, 1344-1356.

Lee, H., Shimizu, Y., Masuda, T., & Uleman, J.S. (2017). Cultural differences in spontaneous trait and situation inferences. *Journal of Cross-Cultural Psychology, 48*, 627-643.

Libby, L.K., & Eibach, R.P. (2011). Visual perspective in mental imagery: A representational tool that functions in judgment, emotion, and self-insight. In M.P. Zanna & J.M. Olson (Eds.), *Advances in Experimental Social Psychology* (Vol. 44, pp. 185-245). San Diego, CS: Academic Press.

Maass, A. (1999). Linguistic intergroup bias: Stereotype perpetuation through language. In Zanna, M. P (Ed.), *Advances in Experimental Social Psychology* (Vol. 31, pp. 79-121). San Diego, CA: Academic Press.

Maass, A., & Arcuri, L. (1992). The role of language in the persistence of stereotypes. In G. R. Semin & K. Fiedler (Eds.), *Language, Interaction and Social Cognition* (pp. 129-143). Newbury Park, CA: Sage.

Maass, A., & Arcuri, L. (1996). Language and stereotyping. In C. N. Macrae, C. Stangor, & M. Hewstone (Eds.), *Stereotypes and Stereotyping* (pp. 193-226). New York: Guilford Press.

Maass, A., Cadinu, M., Boni, M., & Borini, C. (2005). Converting verbs into adjectives: Asymmetrical memory distortions for stereotypic and counterstereotype information. *Group Processes and Intergroup Relations, 8*, 271-290.

Maass, A., Cadinu, M., Taroni, M., & Masserini, M. (2006). The induction-deduction asymmetry: Fact or artifact? *Social Cognition, 24*, 74-109.

Maass, A., Carnaghi, A., & Rakic, T. (2015). Essentialism in language: Plagiarizing David Hamilton. In S.J. Stroessner & J.W. Sherman (Eds.), *Social Perception: From Individuals to Groups* (pp. 212-228). New York: Psychology Press.

Maass, A., Ceccarelli, R., & Rudin, S. (1996). Linguistic intergroup bias: Evidence for in-group-protective motivation. *Journal of Personality and Social Psychology, 71*, 512-526.

Maass, A., Colombo, A., Colombo, A., & Sherman, S.J. (2001). Inferring traits from behaviors and behaviors from traits: The induction-deduction asymmetry. *Journal of Personality and Social Psychology, 81*, 391-404.

Maass, A. Corvino, P., & Arcuri, L. (1994). Linguistic intergroup bias and the mass media. *Revue de Psychologie Sociale*, 1, 31-43.

Maass, A., Karasawa, M., Politi, F., & Suga, S. (2006). Do verbs and adjectives play different roles in different cultures? A cross-linguistic

analysis of person representation. *Journal of Personality and Social Psychology, 90,* 734-750.

Maass, M., Milesi, A. Zabbini, S., & Stahlberg, D. (1995). Linguistic intergroup bias: Differential expectancies or ingroup protection? *Journal of Personality and Social Psychology, 68*, 116-126.

Maass, A., Salvi, D., Arcuri, L., & Semin, G. (1989). Language use in intergroup contexts: The linguistic intergroup bias. *Journal of Personality and Social Psychology, 57*, 981-993.

Mastro, D., Tukachinsky, R., Behm-Morawitz, E., & Blecha, E. (2014). News coverage of immigration: The influence of exposure to linguistic bias in the news on consumer's racial/ethnic cognitions. *Communication Quarterly*, *62*, 135-154.

Markus, H.R., & Kitayama, S. (1991). Culture and the self: Implications for cognition, emotion, and motivation. *Psychological Review, 98*, 224-253.

Masuda, T., & Nisbett, R.E. (2001). Attending holistically versus analytically: Comparing the context sensitivity of Japanese and Americans. *Journal of Personality and Social Psychology, 81,* 922-934.

Na, J., & Kitayama, S. (2011). Spontaneous trait inference is culture-specific: Behavioral and neural evidence. *Psychological Science, 22*, 1025-1032.

Newman, L.S., & Marsden III, A.D. (2023). Around the world in 80 milliseconds (or less): Spontaneous trait inference across cultures. In E. Balcetis & G.B. Moskowitz (Eds.), *The Handbook of Impression Formation: A Social Psychological Approach* (pp. 324-347). New York: Routledge.

Nisbett, R.E., Peng, K.P., Choi, I., & Norenzayan, A. (2001). Culture and systems of thought: Holistic versus analytic cognition. *Psychological Review, 108*, 291-310.

Otten, S., & Moskowitz, G.B. (000). Evidence for implicit evaluative ingroup bias: Affect biased spontaneous trait inference in a minimal group paradigm. *Journal of Experimental Social Psychology, 36,* 77-89.

Ramos, T., Garcia-Marques, L., Hamilton, D.L., Ferreira, M.B., & Van Acker, K. (2012). What I

infer depends on who you are: The influence of stereotypes on trait and situational

spontaneous inferences. *Journal of Experimental Social Psychology, 48,* 1247-1256. Semin, G., & Fiedler, K. (1988). The cognitive functions of linguistic categories in describing persons: Social cognition and language. *Journal of Personality and Social Psychology, 54*, 558-568.

Semin, G., & Fiedler, K. (1991). The linguistic category model, its bases

applications and range. In W. Stroebe & M. Hewstone (Eds.), *European Review of Social Psychology* (Vol. 2, pp. 1-50). Chichester: Wiley.

Semin, G., & Fiedler, K. (Eds.) (1992). *Language, Interaction, & Social Cognition.* London: Sage.

Semin, G., de Montes, L.G., & Valencia, J.F. (2003). Communication constraints on the linguistic intergroup bias. *Journal of Experimental Social Psychology, 39*, 142-148.

Shimizu, Y., Lee, H., & Uleman, J.S. (2017). Culture as automatic processes for making meaning: Spontaneous trait inferences. *Journal of Experimental Social Psychology, 69,* 79-85.

Shulman, J.L., Collins, K.A., & Clement, R. (2011). In consideration of social context: Re-examining the linguistic bias paradigm. *Journal of International and Intercultural Communication, 4,* 310-332.

Srull, T.K., & Wyer, R.S., Jr. (1989). Person memory and judgment. *Psychological Review, 96*, 58-83.

Tajfel, H. (1970). Experiments in intergroup discrimination. *Scientific American, 223*, 96-102.

Tajfel, H., Billig, M.G., Bundy, R.P., & Flament, C. (1971). Social categorization and intergroup behavior. *European Journal of Social Psychology, 1*, 149-177.

Tincher, M.M., Lebois, L.A., & Barsalou, L.W. (2016). Mindful attention reduces linguistic intergroup bias. *Mindfulness, 7*, 349-360.

Todd, A.R., Molden, D.C., Ham, J., & Vonk, R. The automatic and co-occurring activation of multiple social inferences. *Journal of Experimental Social Psychology, 47,* 37-49.

Todorov, A., & Uleman, J.S. (2002). Spontaneous trait inferences are bound to actors' faces: Evidence from a false recognition paradigm. *Journal of Personality and Social Psychology, 83,* 1051-1065.

Todorov, A., & Uleman, J.S. (2003). The efficiency of binding spontaneous trait inferences to actors' faces. *Journal of Experimental Social Psychology, 39,* 549-562.

Todorov, A., & Uleman, J.S. (2004). The person reference process in spontaneous trait inferences. *Journal of Personality and Social Psychology, 87,* 482-493.

Triandis, H.C. (1989). The self and social behavior in differing cultural contexts. *Psychological Review, 96*, 506-520.

Trope, Y., & Liberman, N. (2010). Construal level theory of psychological distance. *Psychological Review, 117*, 440-463.

Uleman, J.S. (1989). A framework for thinking intentionally about unintended thought. In J.S. Uleman & J.A. Bargh (Eds.), *Unintended*

Thought (pp. 425-449). New York, NY: Guilford Press.

Uleman, J.S. (1999). Spontaneous versus intentional inferences in impression formation. In S. Chaiken & Y. Trope (Eds.), *Dual Process Theories in Social Psychology* (pp. 141-160). New York, NY: Guilford Press.

Uleman, J.S., Hon, A., Roman, R., & Moskowitz, G.B. (1996). On-line evidence for spontaneous trait inferences at encoding. *Personality and Social Psychology Bulletin, 22*, 377-394.

Uleman, J.S., Newman, L.S., & Moskowitz, G.B. (1996). People as flexible interpreters: Evidence and issues from spontaneous trait inference. In M.P. Zanna (Ed.), *Advances in Experimental Social Psychology* (Vol. 28, pp. 211-279). San Diego, CA: Academic Press.

Uleman, J.S., Saribay, S.D., & Gonzalez, C.M. (2008). Spontaneous inferences, implicit impressions, and implicit theories. *Annual Review of Psychology, 59*, 329-360.

Vallacher, R.R., & Wegner, D.M. (1987). What do people think they're doing? Action identification and human behavior. *Psychological Review, 94*, 3-15.

Vallacher, R.R., & Wegner, D.M. (1989). Levels of personal agency: Individual variation in action identification. *Journal of Personality and Social Psychology, 57*, 66—671.

Wigboldus, D.H.J., Semin, G.R., & Spears, R. (2000). How do we communicate stereotypes? Linguistic bases and inferential consequences. *Journal of Personality and Social Psychology, 78*, 5-18.

Winter, L., & Uleman, J.S. (1984). When are social judgments made? Evidence for the spontaneousness of trait inferences. *Journal of Personality and Social Psychology, 47,* 237-252.

Zarate, M.A., Uleman, J.S., & Voils, C.I. (2001). Effects of culture and processing goals on the activation and binding of trait concepts. *Social Cognition, 19*, 295-323.

14. Where everything begins: Some hints about the origin of intergroup attitudes among children

Luigi Castelli and Luciana Carraro
University of Padova, Italy

One of the most fascinating endeavors in social psychology is the study of the interplay between cognitive and motivational factors, an interaction that is also critical for the understanding of the development of intergroup attitudes. Even the most robust cognitive mechanisms appear to be most often shaped by motivational forces. An illustrative example is given by illusory correlation phenomena. Indeed, Hamilton and Gifford (1976) have nicely shown that basic cognitive processes could determine the formation of strong stereotypes by simply manipulating the relative frequency of the provided information about social groups. People tend to overestimate the instances of shared infrequency and this may (at least partially) account for the negative attitudes towards minority groups. This was a ground-braking discovery indicating that cognitive processes can deeply affect the perception of our social world, above and beyond the actual characteristics of such world, but this was only part of the story. Importantly, Schaller and Maass (1989) demonstrated that illusory correlation is not impervious to motivational influences and, indeed, when one's own group membership is involved, ingroup-bias appears to largely override the typical illusory correlation effect. This line of research elegantly illustrates the need to consider both cognitive and motivational dynamics when trying to understand the processes involved in the formation of intergroup attitudes. In this chapter, we will first discuss some evidence about the responses of infants and toddlers when faced with

racial ingroup vs outgroup members, arguing that the color of the skin is not a major driver of social behaviors. Next, we will focus on the subsequent developmental phases, from about three years of age, in which various forms of ingroup bias tend to emerge, stressing, however, how such bias is largely shaped by motivational factors and by the perception of the normative context. From this perspective, racial biases would not be conceived as an inevitable and intractable phenomenon, but mainly as the result of the social environment in which children are embedded.

The formation of intergroup attitudes

Not surprisingly, the study about the formation of intergroup attitudes may be maximally informed by research with children. The literature of the last decades about children's intergroup attitudes can be examined as an attempt to pinpoint how cognitive development and social factors (e.g., group identity, threat, status, influence of significant adults) may intervene in shaping such attitudes. The influential socio-cognitive theory proposed by Aboud (1988; Aboud & Doyle, 1996) – adopting a neo-Piagetian perspective – is based on the idea that different psychological processes characterize different developmental phases, and the nature of children's responses are directly influenced by which process is prevalent at each particular stage (Aboud, 2008). As for the very early stage of intergroup attitudes formation, the theory assumes that children's responses are essentially dominated by affective experiences. In particular, very young children would soon start responding differently towards different social groups mainly as a function of their familiarity: Unknown groups give raise to feelings of threat whereas an attachment is developed towards the more "secure" familiar groups. Some data with children in their first months of life (but not newborns; Kelly et al., 2005) would seem to be consistent with this view, in that the analysis of infants' looking patterns indicates a visual preference for own-race faces as compared to other-race faces (Bar-Haim et al., 2006; Kelly et al., 2007; see also Anzures et al., 2013). However, it is not yet unambiguously clear whether visual preferences can be directly equated to positive attitudes, namely to more positive affective responses and approach tendencies (see Rhodes, 2020, for a detailed discussion). Indeed, empirical findings seem to suggest that they should be better conceived as separated aspects. For instance, Kinzler and Spelke (2011) have nicely demonstrated that while 5-year-old White children express strong ingroup bias in an interracial

context, 10-month-old White infants appear to be equally likely to take toys from either a White or a Black adult. In sum, preverbal children in their studies provided color-blind responses suggesting that, overall, at early developmental stages infants can easily distinguish individuals as a function of the color of their skin, but infants' social behavior is not necessarily affected by the color of the skin of the potential interaction partner (Liberman et al., 2017). We recently replicated these findings with a sample of White Italian children aged between 1 and 2 years (Castelli & Carraro, 2020). Each participant was shown two adult actors – one White and one Black – who started eating a piece of bread (or a large candy) and then offered the remaining part to the participant. The key dependent variable was the identity of the adult from whom the child decided to take the bread (or the candy). Consistent with the findings obtained by Kinzler and Spelke (2011), children during their second year of life were equally likely to take the food from either the White or the Black adult. Importantly, with the administration of the same experimental procedure to children between 3 and 4 years, a significant and strong (d = 1.01) preference for the food handled and eaten by the White actor emerged (Castelli & Carraro, 2020). Hence, social discrimination based on the color of the skin does not seem to appear before 3 years of age and it can thus be hardly considered as an innate (pro-ingroup) phenomenon (Rhodes, 2020). The theoretical proposal put forward by Drew Nesdale (i.e., Social Identity Development Theory, SIDT; Nesdale, 1999, 2004) makes exactly this prediction, assuming that, prior to 2-3 years of age, children go through an undifferentiated phase in which racial cues are relatively irrelevant for children who do not use them for navigating their social world. In contrast, it is likely that, as age increases, thanks to cognitive maturation, children become more sensitive to the influence of significant others and to social norms which, in turn, Not surprisingly, the study about the formation of intergroup attitudes may be maximally informed by research with children. The literature of the last decades about children's intergroup attitudes can be examined as an attempt to pinpoint how cognitive development and social factors (e.g., group identity, threat, status, influence of significant adults) may intervene in shaping such attitudes. The influential socio-cognitive theory proposed by Aboud (1988; Aboud & Doyle, 1996) – adopting a neo-Piagetian perspective – is based on the idea that different psychological processes characterize different developmental phases, and the nature of children's responses are directly influenced by which process is prevalent at each particular stage (Aboud,

2008). As for the very early stage of intergroup attitudes formation, the theory assumes that children's responses are essentially dominated by affective experiences. In particular, very young children would soon start responding differently towards different social groups mainly as a function of their familiarity: Unknown groups give raise to feelings of threat whereas an attachment is developed towards the more "secure" familiar groups. Some data with children in their first months of life (but not newborns; Kelly et al., 2005) would seem to be consistent with this view, in that the analysis of infants' looking patterns indicates a visual preference for own-race faces as compared to other-race faces (Bar-Haim et al., 2006; Kelly et al., 2007; see also Anzures et al., 2013). However, it is not yet unambiguously clear whether visual preferences can be directly equated to positive attitudes, namely to more positive affective responses and approach tendencies (see Rhodes, 2020, for a detailed discussion). Indeed, empirical findings seem to suggest that they should be better conceived as separated aspects. For instance, Kinzler and Spelke (2011) have nicely demonstrated that while 5-year-old White children express strong ingroup bias in an interracial context, 10-month-old White infants appear to be equally likely to take toys from either a White or a Black adult. In sum, preverbal children in their studies provided color-blind responses suggesting that, overall, at early developmental stages infants can easily distinguish individuals as a function of the color of their skin, but infants' social behavior is not necessarily affected by the color of the skin of the potential interaction partner (Liberman et al., 2017). We recently replicated these findings with a sample of White Italian children aged between 1 and 2 years (Castelli & Carraro, 2020). Each participant was shown two adult actors – one White and one Black – who started eating a piece of bread (or a large candy) and then offered the remaining part to the participant. The key dependent variable was the identity of the adult from whom the child decided to take the bread (or the candy). Consistent with the findings obtained by Kinzler and Spelke (2011), children during their second year of life were equally likely to take the food from either the White or the Black adult. Importantly, with the administration of the same experimental procedure to children between 3 and 4 years, a significant and strong (d = 1.01) preference for the food handled and eaten by the White actor emerged (Castelli & Carraro, 2020). Hence, social discrimination based on the color of the skin does not seem to appear before 3 years of age and it can thus be hardly considered as an innate (pro-ingroup) phenomenon (Rhodes, 2020). The theoretical proposal put forward

by Drew Nesdale (i.e., Social Identity Development Theory, SIDT; Nesdale, 1999, 2004) makes exactly this prediction, assuming that, prior to 2-3 years of age, children go through an undifferentiated phase in which racial cues are relatively irrelevant for children who do not use them for navigating their social world. In contrast, it is likely that, as age increases, thanks to cognitive maturation, children become more sensitive to the influence of significant others and to social norms which, in turn, leads to a differential treatment of social groups.leads to a differential treatment of social groups.

Familiarity-based explanations of racial ingroup bias among very young children cannot account for a variety of different research findings. First, as discussed above, although familiarity affects visual preferences in a pro-ingroup direction (Bar-Haim et al., 2006; Kelly et al., 2007), no similar pattern can be observed in relation to actual discrimination among children under 3 years of age (Castelli & Carraro, 2020; Kinzler & Spelke, 2011). Second, when socially meaningful attitudes (e.g., differential evaluative judgments) start to emerge at about age 3, their trajectories appear to differ for majority and minority group members. Hence, even mere perceptual processes, which at that stage enable a better assessment of the similarity between oneself and others, can hardly provide a comprehensive account of empirical findings. Indeed, while majority group members develop a preference for the ingroup (e.g., Aboud, 1988; Castelli et al., 2007; Raabe & Beelmann, 2011), members of low-status groups often display a preference for the outgroups that, in the specific social context, are associated with a higher social status (e.g., Dunham et al., 2007, 2013; Newheiser & Olson, 2012). In this regard, clear-cut findings emerge from studies conducted in South Africa, in which Black children – simultaneously representing the numeric majority but also the minority group in relation to social status - show a preference for White rather than Black targets (Dunham et al., 2014; Newheiser et al., 2014; Shutts et al., 2011). This supports the idea that it is indeed the understanding of the relative position occupied by one's own group in the social hierarchy that contributed in shaping early intergroup attitudes, and not just a feeling of familiarity or the preference preference for individuals who are appraised as similar rather than dissimilar to oneself.

Conceptually consistent findings emerge from studies relying on the so-called minimal group paradigm in which participants are assigned to a specific group on the basis of trivial criteria (Tajfel et al., 1971). In this regard, Nesdale and Flesser (2001) assigned 5- and 8-year-old children to

a minimal group and manipulated the status of the ingroup that could be either higher or lower as compared to that of the outgroup. Importantly, ingroup bias was significantly reduced for participants who believed to belong to a low-status group indicating that children are very responsive the perceived value associated with social groups, above and beyond the mere ingroup-outgroup distinction. Notably, the information about the status of the ingroup had also an impact on the motivation to either remain in the group or to change one's own membership (see also Nesdale et al., 2004). The results reported by Nesdale and Flesser (2001) have been recently replicated by Castelli et al. (2021) with both explicit and implicit (i.e., Child-IAT) attitude measures, suggesting that the knowledge about the relative status of the groups does not only affect deliberate responses (i.e., a thoughtful combination of all the provided information about the groups) but also children's more spontaneous and less controlled responses.

Overall, it appears that children from about 3 years of age are very attentive to the characteristics and value that others associate with their own group, and structure their judgments and behaviors accordingly. The work by Nesdale, Maass, and colleagues (Nesdale, 2004; Nesdale & Flesser, 2001; Nesdale et al., 2004) directly links these phenomena to the premises of social identity theory (Tajfel & Turner, 1979) and, more specifically, to the idea that children derive part of their self-worth from their group memberships and, therefore, they strive to belong to the groups that are more highly regarded in a given context. Importantly, although children from 4 to 6-7 years tend to strongly favor their own groups over other groups - at least in the case of majority group members – the attitudes towards the outgroups are also typically positive (or neutral; see Cameron et al., 2001). Hence, at this developmental stage, it is mainly a matter of a differential positivity associated to different groups with no particular strong evidence of a pervasive active derogation of outgroups. This, however, does not imply that children do not simply pay attention to outgroups and that the attitudes towards outgroups do not reflect meaningful individual differences. For instance, it has been shown that the attitudes towards the racial ingroup can be negatively correlated with those towards a racial outgroup (Aboud, 2003), suggesting that children likely carry out an intergroup comparison and that a strong attachment to the ingroup may be one determinant of less positive evaluations of the outgroup (explaining around 15% of the variance). In the work by Aboud (2003) such correlation, however, was no longer observable in a sample

of respondents with a relatively strong familiarity with the outgroup (i.e., in a mixed-race school), stressing one more time the importance of intergroup contact (Tropp & Prevost, 2008). When little direct knowledge is present, ingroup attachment may become the primary basis for structuring the attitude towards the outgroup, whereas it does no longer play a significant role when personal experiences are allowed and a direct knowledge about outgroup members arises. Similar findings have been also reported in the Italian context (Carraro & Castelli, 2015). Indeed, White children - aged between 3 and 5 years – with a previous history of contact with Black but not with Asian peers displayed the aforementioned negative correlation between the evaluation of the ingroup and the outgroup only in relation to the perception of Asian targets (Carraro & Castelli, 2015).

A key theoretical and practical question is how intergroup attitudes develop afterwards. The socio-cognitive theory from Aboud (1988) and Nesdale's social identity development theory (SIDT; Nesdale, 1999, 2004) make here somehow different predictions. According to the former theory, ingroup bias is expected to decline due to the increasing cognitive abilities of the children who - around 7 years of age – become gradually more capable of taking the perspective of others, thus reducing their reliance on categorical distinctions in favor of an appreciation of individual characteristics. SIDT, in contrast, does not make this straightforward prediction and posits that intragroup processes and intergroup dynamics are major determinants of whether ingroup bias will either increase or decrease. In a seminal study, Nesdale, Maass, Durkin, and Griffiths (2005) have outlined the key role of group norms and threat. In particular, the authors manipulated whether the ingroup was described as possessing an inclusion or exclusion norm (i.e., the extent to which the ingroup was overall willing to include/exclude in their activities other children who happened to be different) and whether the outgroup could represent a threat for the status of the ingroup. Results indicated that both the exclusion norm and the perceived threat led to more negative attitudes towards the outgroup and, more specifically, that when the ingroup was believed to embrace an exclusion norm blatant prejudice was more likely to emerge (i.e., evaluations were negative in absolute terms). Hence, although meta-analyses indicate that ingroup bias tends to decline after age 7 (see Raabe & Beelmann, 2011), the developmental pattern is neither rigid nor ubiquitous and socio-motivational factors may strongly influence its trajectory. This assumption is further supported by evidence

indicating that children's emotional empathy is indeed a significant predictor of more positive attitudes towards the outgroup – in line with the prediction of a sociocognitive theory – but such relationship is disrupted when the ingroup has an exclusion norm (Nesdale et al., 2005). Hence, cognitive development generally allows children to better take the perspective of others and to understand their feelings and internal states, but the motivation to follow the norms of the ingroup may have hurtful effects when such norms legitimate (or even foster) outgroup derogation. Again, a strict interplay between cognitive and motivational factors can be observed.

One overall message stemming from these studies is that children's attitudes are attuned to the perceived beliefs of the significant others around them. As the child grows up, the influence of peers becomes maximally relevant (see Zingora et al., 2020). Among younger children, however, the influence of the primary caregivers (e.g., parents and teachers) is likely to be predominant. It is noteworthy that often parents appear to be quite reticent about discussing racial issues with their children (Vittrup, 2018; Vittrup & Holden, 2011). The prevalent approach of parents belonging to majority groups is to adopt a color-blind perspective when it comes to have conversations with children that might have race-related implications. Although adults are unlikely to provide relevant information about race-related issues through a verbal channel, an increasing literature suggests that during the early phases of intergroup attitudes formation nonverbal signals may play a critical role (Castelli et al., 2008; Skinner et al., 2017; 2020). Indeed, young children appear to grasp the subtle nonverbal behaviors of adults during interracial interactions and, as a consequence, they model their own attitudes as a function of such observed nonverbal behaviors. These findings can also be interpreted at light of the different ways in which information about group norms can be acquired and expressed. On the one hand, injunctive norms can be verbally transmitted, communicating what is either approved or disapproved, as in the case of blatant exclusion norms. On the other hand, descriptive norms are a potent vehicle of cultural transmission. Through the observation of what others do (e.g., increasing the interpersonal distance from members of stigmatized groups), children may likewise build up representations about whether negative (vs. positive) behaviors towards outgroup members can be considered as normatively appropriate.

To conclude, cognitive constraints inevitably set the boundaries in relation to the characteristics of the intergroup attitudes of the devel-

oping child, but social factors are likely to be far more relevant. In this regard, the literature is clear in highlighting the importance to set up social environments that enable meaningful interactions between different ethnic groups (Tropp & Prevost, 2008), to promote color conscious approaches and discussions about race-related issues with children (Perry et al., 2024), and to support inclusive norms both through structured educational interventions and one's subtle everyday behaviors. In order to achieve more egalitarian societies, this stresses the huge responsibilities of parents, educators, and all those who interact and take care of child, always keeping in mind the idea that young children are not inherently predisposed to racism.

References

Aboud, F. E. (1988). Children and prejudice. New York: Blackwell.

Aboud, F. E. (2003). The formation of in-group favoritism and outgroup prejudice in young children: Are they distinct attitudes? Developmental Psychology, 39, 48-60.

Aboud, F. E. (2003). A social-cognitive developmental theory of prejudice. In S. M. Quintana & C. McKown (Eds), Handbook of race, racism, and the developing child (pp. 55- 71). Hoboken, NJ: Wiley.

Aboud, F. E., & Doyle, A. B. (1996). Parental and peer influences on children's racial attitudes. International Journal of Intercultural Relations, 20, 371-383.

Anzures, G., Quinn, P. C., Pascalis, O., Slater, A. M., Tanaka, J. W., & Lee, K. (2013). Developmental origins of the Other-Race Effect. Current Directions in Psychological Science, 22, 173–178.

Bar-Haim, Y., Ziv, T., Lamy, D., & Hodes, R. M. (2006). Nature and nurture in own-race face processing. Psychological Science, 17, 159–163.

Cameron, J. A., Alvarez, J. M., Ruble, D. N., & Fuligni, A. J. (2003). Children's lay theories about ingroups and outgroups: Reconceptualizing research on prejudice. Personality and Social Psychology Review, 5, 118-128.

Carraro, L., & Castelli, L. (2015). On the generality of children's racial attitudes across target groups. Psicologia sociale, 10, 71-80.

Castelli, L., & Carraro, L. (2020). No evidence of racial discrimination among toddlers. Psicologia Sociale, 15, 285–292.

Castelli, L., Carraro, L. & Valmori, A. (2021). Group status rapidly shapes preschoolers' social judgments in minimal group settings. Journal of

Experimental Child Psychology, 206, 105102.

Castelli, L., De Amicis, L., & Sherman, S. J. (2007). The loyal member effect: On the preference for ingroup members who engage in exclusive relations with the ingroup. Developmental Psychology, 43, 1347-1359.

Castelli, L., De Dea, C., & Nesdale, D. (2008). Learning social attitudes: Children's sensitivity to the verbal and nonverbal behaviors of adult models during interracial interactions. Personality and Social Psychology Bulletin, 34, 223-237.

Dunham, Y., Baron, A. S., & Banaji, M. R. (2007). Children and social groups: A developmental analysis of implicit consistency in Hispanic Americans. Self and Identity, 6, 238-255.

Dunham, Y., Newheiser, A. K., Hoosain, L., Merrill, A., & Olson, K. R. (2014). From a different vantage: Intergroup attitudes among children from low- and intermediate-status racial groups. Social Cognition, 32, 1-21.

Hamilton, D. L., & Gifford, R. K. (1976). Illusory correlation in interpersonal perception: A cognitive basis of stereotypic judgments. Journal of Experimental Social Psychology, 12, 392-407.

Kelly, D. J., Quinn, P. C., Slater, A. M., Lee, K., Ge, L., & Pascalis, O. (2007). The other-race effect develops during infancy. Psychological Science, 18, 1084-1089.

Kelly, D. J., Liu, S., Ge, L., Quinn, P. C., Slater, A. M., Lee, K., Liu, Q., & Pascalis, O. (2007). Cross-race preferences for same-race faces extend beyond the African versus Caucasian contrast in 3-month-old infants. Infancy, 11, 87-95.

Kinzler, K. D., & Spelke, E. S. (2011). Do infants show social preferences for people differing in race? Cognition, 119, 1-9.

Liberman, Z., Woodward, A. L., & Kinzler, K. D. (2017). The origins of social categorization. Trends in Cognitive Sciences, 21, 556-568.

Nesdale, D. (1999). Social identity and ethnic prejudice in children. In P. Martin & W. Noble (Eds.), Psychology and society (pp. 92-110). Brisbane: Australian Academic Press.

Nesdale, D. (2004). Social identity processes and children's ethnic prejudice. In M. Bennett & F. Sani (Eds.), The development of the social self (pp. 219-246). London: Psychology Press.

Nesdale, D., Durkin, K., Maass, A., & Griffiths, J. (2004). Group status, outgroup ethnicity and children's ethnic attitudes. Journal of Applied Developmental Psychology, 25, 237-251.

Nesdale, D., & Flesser, D. (2001). Social identity and the development of children's group attitudes. Child Development, 72, 506-517.

Nesdale, D., Griffith, J., Durkin, K., & Maass, A. (2005). Empathy, group norms and children's ethnic attitudes. Journal of Applied Developmental Psychology, 26, 623-637.

Nesdale, D., Maass, A., Durkin, K., & Griffiths, J. (2005). Group norms, threat, and children's racial prejudice. Child Development, 76, 652-663.

Newheiser, A. K., Dunham, Y., Merrill, A., Hoosain, L., & Olson, K. R. (2014). Preference for high status predicts implicit outgroup bias among children from low-status groups. Developmental Psychology, 50, 1081-1090.

Newheiser, A. K., & Olson, K. R (2012). White and Black American children's implicit intergroup bias. Journal of Experimental Social Psychology, 48, 264-270.

Perry, S., Wu, D. J., Abaied, J. L., Skinner-Dorkenoo, A. L., Sanchez, S., Waters, S. F., & Osnaya, A. (2024). White parents' racial socialization during a guided discussion predicts declines in white children's pro-white biases. Developmental Psychology, 60, 624–636.

Raabe, T., & Beelmann, A. (2011). Development of ethnic, racial, and national prejudice in childhood and adolescence: A multinational meta-analysis of age differences. Child Development, 82, 1715-1737.

Rhodes, M. (2020). Are humans born to hate? Three myths and three developmental lessons about the origins of social categorization and inter-group bias. In J. Decety (Ed.), The social brain: A developmental perspective. MIT Press.

Schaller, M., & Maass, A. (1989). Illusory correlation and social categorization: Toward an integration of motivational and cognitive factors in stereotype formation. Journal of Personality and Social Psychology, 56, 709-721.

Shutts, K., Kinzler, K. D., Katz, R. C., Tredoux, C., & Spelke, E. S. (2011). Race preferences in children: Insights from South Africa. Developmental Science, 14, 1283-1291.

Skinner, A. L., Meltzoff, A. N., & Olson, K. R. (2017). "Catching" social bias: Exposure to biased nonverbal signals creates social biases in preschool children. Psychological Science, 28, 216-224.

Skinner, A. L., Olson, K. R., & Meltzoff, A. N. (2020). Acquiring group bias: Observing other people's nonverbal signals can create social group biases. Journal of Personality and Social Psychology, 119, 824.

Tajfel, H., Billig, M. G., Bundy, R. P., & Flament, C. (1971). Social categorization and intergroup behaviour. European Journal of Social Psychology, 1, 149-178.

Tajfel, H. & Turner, J. (1979). An integrative theory of intergroup conflict. In W. G. Austin & S. Worchel (Eds.), The social psychology of intergroup relations (pp. 33-47). Monterey, CA: Brooks/Cole Publishing.

Tropp, L. R., & Prenovost, M. A. (2008). The role of intergroup contact in predicting children's interethnic attitudes: Evidence from meta-analytic and field studies. In S. R. Levy & M. Killen (Eds.), Intergroup attitudes and relations in childhood through adulthood (pp. 236–248). Oxford University Press.

Vittrup, B. (2018). Color blind or color conscious? White American mothers' approaches to racial socialization. Journal of Family Issues, 39, 668– 692.

Vittrup, B., & Holden, G. W. (2011). Exploring the impact of educational television and parent–child discussions on children's racial attitudes. Analyses of Social Issues and Public Policy, 11, 82-104.

Zingora, T., Stark, T. H., & Flache, A. (2020). Who is most influential? Adolescents' intergroup attitudes and peer influence within a social network. Group Processes & Intergroup Relations, 23, 684-709.

15. Social rationality

Bruno Gabriel Salvador Casara
New York University Abu Dhabi, United Arab Emirates

Rationality refers to being guided by reason or being based on logic The concept of rationality, in this sense, refers to acting rationally if you have a good reason for what you are doing or believing rationally if there is strong evidence to support your belief. How rational are human beings? This question raised fervid debates in science. In Economics, and particularly in the Expected Utility Hypothesis (von Neumann & Morgenstern, 1947), it is assumed that humans are capable of pursuing optimally their subjective-defined goals. After estimating the utility of an outcome, it is hypothesized that humans will choose the alternative with the higher grade of utility (Savage, 1954). For example, an individual will maximize utility as a consumer and profit as a producer. Therefore, humans are viewed as rational agents, characterized by clear preferences and able to evaluate actions by taking into account a huge amount of available information, probabilities of events, and potential costs and benefits, and finally, capable of acting with coherence in the decision of the self-determined best choice of action. Similar ideas are also present and shared in social psychology. Indeed, at least two theories in Social Psychology are strongly in line with the Expected Utility Hypothesis: The Theory of Reasoned Action (Hill et al., 1977), and the Theory of Planned Behavior (Ajzen, 2011).

According to both these theories, individual behavioral intentions are the result of individual attitudes and subjective norms. For example, a student will be more likely to enroll in a Social Psychology course if he or she likes the idea of enrolling in that course (positive attitude) and believes that this behavior will be accepted by his or her significant oth-

ers (subjective norms). Moreover, the Theory of Planned Behavior states that also perceived behavioral control over the behavior will predict the behavioral intention: I may like the idea to sing, and I may believe that people around me will appreciate hearing a song, but I will not sing if I believe that I am out of tune!

The reader now may notice how these theories are in line with the view of humans as rational agents. Indeed, all these theories are based on the idea that behavior is the product of evaluation processes where a potential behavior has to satisfy personal attitudes, norms, and feasibility aspects. However, individuals and groups often act against their interests and perceived norms, for example, people can continue to smoke, even if they do not enjoy it anymore and they feel that it is not a behavior accepted by their group.

The conception of individuals as rational agents was criticized by Herbert Simon (1957) who proposed that individuals rarely are capable of applying processes that maximize expected utility. Otherwise, bounded rationality drives more plausible to a fast and simple process called heuristics. These heuristics provide satisfactory results with a reasonable cognitive effort, even if they are not necessarily the best decisions or the most realistic perceptions. This happens because reality is extremely rich in stimuli, and our cognitive abilities do not allow us to deeply analyze every situation we have to face. So, it is very likely that even Albert Einstein, whose theory was able to predict how gravity works near black holes at the center of our galaxy (Hannam et al., 2022), sometimes locked himself out of his own house. Interestingly, inconsistencies in human reasoning are not only limited due to the boundaries of our cognitive abilities and the propensity of avoiding cognitive efforts. Indeed, people are also motivated to avoid systematic reasoning and to avoid the opportunity to be open to new attitudes. But why people should have such motivations? First of all, people want to be able to interpret the reality that they experienced. While this desire leads us to look for accurate information, this also implies that information that contradicts our worldviews and previous knowledge is destabilizing and creates cognitive dissonance (Festinger, 1962). To avoid losing our tools for interpreting the world, and to maintain coherence in our experiences, we exhibit cognitive processes and behaviors limiting and distorting our information acquisition, elaboration, and interpretation. For example, people tend to avoid information sources that are likely to express counter-attitudinal positions (Cinelli et al., 2021). Finally, like in Aesop's tale about the fox and the grape,

where the fox convinces herself to dislike the grape just because it cannot reach it, the experiences are more likely to be interpreted in order to satisfy self-serving needs (Arkin et al., 1980), meaning that we trick ourselves in order to maintain a more positive picture of ourselves and the situations we live. Furthermore, the understanding of reality is not the only need that characterizes human motivation, we are not alone in our reality, and our interpretation of the world interacts and sometimes clashes with one of other individuals and groups: we are social animals, and finding people we like and being accepted by others is fundamental to our existence (Stillman & Baumeister, 2009; Tajfel et al., 1979). Thus, it may not too surprising to know that the deficiencies in our judgments and decisions are strongly related to the relationships we have with other people and groups (Clark et al., 2019). We can embrace extremely radical beliefs just because some significant others endorse those beliefs. We avoid to processes situations analytically because we may trust the expertise and the goodwill of other persons. We can express attitudes and behaviors endorsed by our group just to display our loyalty. In sum, once again rationality can be neglected because our thoughts are influenced, often without awareness, by our social bonds.

To sum up, the original idea of human nature characterized by perfect rationality has been challenged by these alternative views: the cognitive miser (e.g Kahneman & Tversky, 1973), the consistency seeker (Festinger, 1962), and the group member (Tajfel et al., 1979). These perspectives received fundamental empirical support over the years and the psychological literature demonstrated that our predictions and judgments, and consequently our decisions, are systematically distorted by several cognitive processes. These systematic distortions are called cognitive biases. A bias can be viewed as a deviation from the norm or rationality (Haselton et al., 2005). Other contents that can be viewed as the result of bounded rationality are logical fallacies, which can be defined as assertions that contained arguments in which the premise fails to support the conclusion.

While heuristics, cognitive bias, and logical fallacies are all related to how people reason and take decisions, and they are all processes potentially leading to suboptimal decisions, they also have important differences and are distinct concepts. Indeed, heuristics are actually generally adaptive and help us to make quick and efficient judgments in everyday life. However, while heuristics are not inherently flawed, the following paragraphs will highlight cases in which heuristics lead to errors in judgment. Differently, cognitive biases are systematic errors by definitions.

Thus, unlike cognitive heuristics, cognitive biases are always maladaptive and lead to errors in judgment and decision-making. Finally, logical fallacies, unlike cognitive heuristics and biases, are not related to how we process information, but rather to how we reason and argue.

In the next paragraphs are presented the main heuristics, cognitive bias, and logical fallacies.

Heuristics

Anchoring

People base their estimations based on a first perceived value (Tversky & Kahneman, 1974). The initial value, the anchor, is considered a piece of relevant information in the elaboration of estimates and the correction of this information is usually insufficient (Slovic & Lichtenstein, 1971). Tversky and Kahneman (1974) were the first to propose the anchoring heuristic. In one of their initial experiments, participants were given five seconds to calculate the product of numbers one through eight, presented either in ascending or descending order. As participants didn't have sufficient time to complete the calculation, they made an estimate after performing a few initial multiplications. When the initial multiplication gave a smaller number, the median estimate was 512, whereas it was 2,250 when the initial numbers were larger (The correct answer was 40,320).

Affect heuristic

Objects and events are sentimentally categorized with the quality of "goodness" or "badness". The use of these sentimentalized categories for judgments and decisions is named the affect heuristic (Slovic et al., 2007). In other words, the emotional reactions to risks can predict the estimated likelihood of a risk.

For example, when individuals evaluate the safety of vaccination they may be influenced by the emotion that they have experienced. Picture showing crying infants near a big syringe may trigger emotions of fear in the parents, these emotions are used as information in the judgment of vaccination safety and so it is plausible that parents will overestimate the danger of vaccines.

Availability

The availability heuristic consists of the evaluation of the frequency or the likelihood of an event on the base of the easiness of representing similar events in mind (Tversky & Kahneman, 1974). Specifically, Tversky and Kahneman investigated the availability heuristic in their initial study (1973) by posing the question, "Which is more common in English: words that begin with the letter 'K', or words in which 'K' is the third letter?" The researchers reasoned that participants would more easily recall words that begin with 'K' such as "kangaroo" or "kitchen" than words in which 'K' is the third letter such as "acknowledge" or "ask". Results showed that participants tended to overestimate the frequency of words that begin with 'K' and underestimate the frequency of words in which 'K' is the third letter. In other words, the perceived frequency of an event can be influenced by the easiness of the memory recall of similar events, so the events that are easier to recall are considered more frequent in the real world (Schwarz & Vaughn, 2002). Furthermore, the tasks that are easy to imagine are considered also easier to perform (Sherma et al., 1985).

Representativeness

The heuristic of representativeness (Tversky & Kahneman, 1983) is about associating similar characteristics to events or subjects that are perceived as similar. In a classic experiment of Tversky and Kahneman (1983) undergraduates were given a description of Linda, a feminist, and asked to evaluate the probability of her being a feminist, a bank teller, or both. Probability theory states that the probability of a conjunction (both a feminist and bank teller) cannot be greater than the probability of either constituent alone. However, participants rated the conjunction as more probable than the bank teller alone. Thus, the representativeness heuristic is used when individuals judge someone based on a prototype. Stereotypes are an example of the representativeness heuristic application; two different individuals can be judged as cold and competent because both are representative of the Jewish ethnic group, which is often stereotyped in that way (Cuddy et al., 2008).

Biases

Confirmatory bias

The tendency of recalling, interpreting, and searching information that is in line with pre-existing attitudes and opinions is named confirmatory bias (Nickerson, 1998). Furthermore, this bias can compromise the ability of an individual to produce and rationally evaluate counterarguments to his or her initial idea (Baron, 1995). Confirmatory biases can be expressed in different phases of information retrieving and processing, for example people are more likely to perceive pro-attitudinal information, and therefore are more likely to select it. Consequently, this information is more likely to be remembered and then used in future situations. In a Study of Salvador Casara et al. (2019), participants had to seek for vaccines-related information in an internet-search context, the results showed that participants' behavior was consistent with confirmatory bias: participants used keywords, put more attention to, and decided to read, web articles consonant with their pre-existing attitudes towards vaccines. This bias can also explain how discredited beliefs are maintained, indeed, the debriefing of fake information is less effective on people who believe in it (Ross, et al., 1975).

Conformism

People tend to behave in order to match social norms, attitudes, and behaviors that are perceived as accepted and shared by others. Studies on the influence of the majority showed that people deliberately and frequently accept distrust and abjure their judgments and perceptions (even when they are correct) when they believe that the majority unanimously holds different positions (Asch, 1956). In the Asch's experiment, participants were asked to judge the length of lines on a set of cards, with one line being the "standard" and the others being comparison lines. The task was easy and unambiguous, and the correct answer was obvious. However, the participants were tested in groups, with confederates of the experimenter deliberately giving incorrect answers. Asch found that around 75% of participants conformed to the group at least once, giving an incorrect answer.
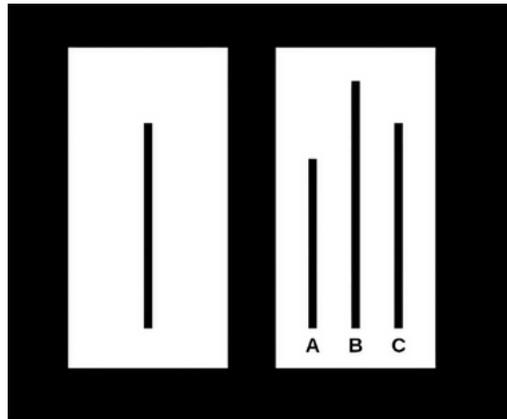
Figure 1. Example of cards used in Asch's experiments.

Conjunction fallacy

The conjunction fallacy consists of the overestimation of co-occurrent events (Tversky & Kahneman, 1983). For example, specific situations (e.g., the co-occurrence of a symptom and a vaccination) can be considered more probable than generic ones (e.g., the presence of a symptom in general), because the specific situation is more representative of a perceived prototypical situation, conjunction fallacy can be considered as an outcome of representative heuristic.

Dunning-Kruger effect

The Dunning-Kruger effect (Kruger & Dunning, 1999) is a cognitive bias that consists of the tendency of considering themselves as more competent than they are when they have low ability in the skill evaluated. This effect seems due to a metacognitive lack of recognizing the own ineptitude. On the other hand, people with high ability tend to underestimate themselves as they erroneously believe that a task easy for them is easy even for others.

Fundamental attribution error

Fundamental attribution error (known also as the actor-observer bias, Kanouse et al., 1972, or as correspondence bias, Gilbert & Malone, 1995) consists of a tendency of judging others' behaviors using dispositional characteristics rather than situational ones. First evidence for this bias was provided by an experiment conducted by Jones and Harris (1967). In

this experiment participants (college students) were asked to read essays written by fellow students. The essays were either pro- or anti-Castro, and the participants were asked to rate the authors' true attitudes towards Fidel Castro. Before reading the essays, half of the participants were told that the authors had freely chosen their position on Castro, while the other half were told that the authors had been assigned their position and had no choice in the matter. The results showed that participants rated the authors' attitudes towards Castro as more consistent with the position they had taken in their essays, even when they were told that the authors had no choice in their position. This suggests that the participants were using dispositional explanations for the authors' attitudes rather than situational explanations, even when situational factors were clearly present.

The likelihood of falling for this bias seems to be predicted by several factors; one of these is culture. Indeed, individualistic cultures (such as western cultures) are more prone to this bias in comparison with collective cultures (such as eastern cultures; Miller, 1984). Furthermore, usually, we are more aware of the actor of behavior rather than the situation in which the behavior emerges, thus, the salience of the actor may drive the fundamental attribution error (Lassiter et al., 2002).

Hostile attribution bias

The tendency in attributing hostile intentions to others is named hostile attribution bias (Nasby, et al., 1980). A person sensitive to this bias will interpret ambiguous social situations in a hostile way, for example, s/he may attribute derogatory goals to laughing people, as s/he will perceive her/himself as the target of the laughs.

Ingroup bias

People tend to favor the ingroup members, even when groups are built using random or trivial criteria (Tajfel, 1970). In particular, Tajfel, using the minimal group paradigm, was able to demonstrate that categorizing people in different group is a sufficient condition in order to let ingroup bias to emerge. The minimal group paradigm involves creating groups based on arbitrary criteria, such as a random assignment to a "blue" or "green" group or sorting participants based on their preference for abstract paintings. Participants are then informed of their group membership and given opportunities to distribute rewards or punishments to

individuals in their own group or in the other group. The main result of studies using this paradigm is that participants tend to allocate more resources and rewards to members of their own group, even when it comes at a cost to members of other groups.

Further studies highlighted that ingroup favoritism affects judgments and behaviors in several different ways, for example, studies suggest that the information provided by ingroup members is more easily accepted, whereas information provided by outgroup members neglected or challenged, and collective actions are more easily taken when people have a stronger identification with the ingroup (Van Zomeren, 2013).

Loss aversion

Is it better to gain 200 dollars or to avoid a fine of the same amount of money? If you prefer the second option, you are in good company. The studies conducted by Kahneman and Tversky (e.g. Kahneman & Tversky 2013) demonstrated that people generally are preferred to avoid losses rather than achieve gain, even when the amount of potential loss or gain is the same. This bias has several practical implications: for example, loss aversion can influence investment decisions, causing people to hold onto losing investments longer than they should or avoiding risks altogether. Moreover, people may fear losing their loved ones and often engage in behaviors that protect their relationships and avoid losses rather than focus on build new and potentially more satisfying romantic relationships. In the education context, students may be more motivated to avoid failing a test or receiving a low grade than they are to strive for a high grade.
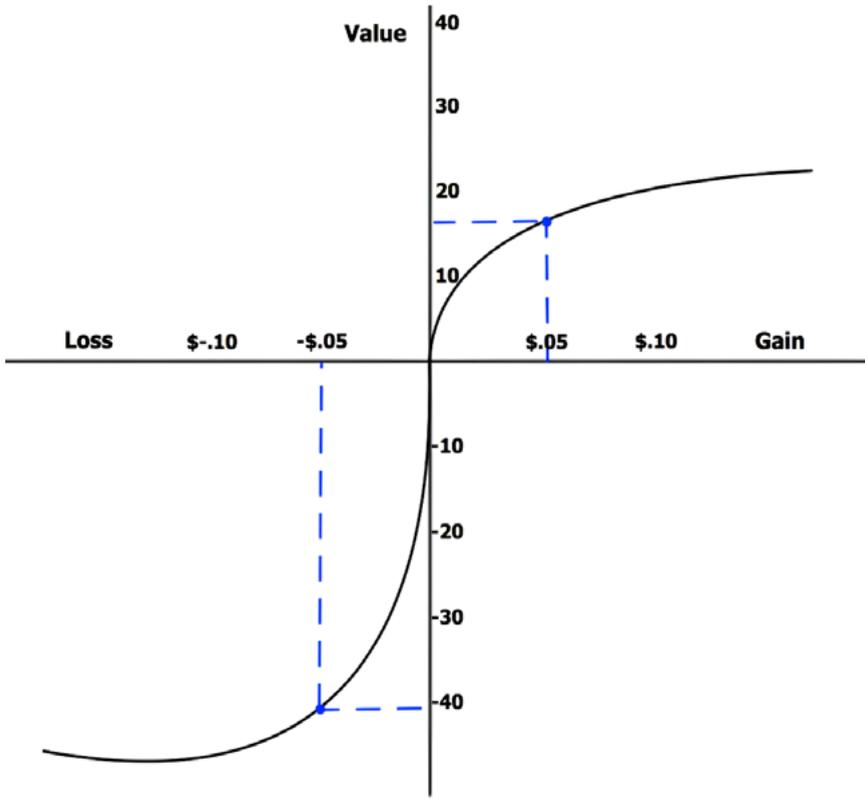
Figure 3. Perceived value when it is framed as a gain vs. loss.

Unrealistic Optimism

People tend to expect that negative events are more likely to happen for others than themselves, and vice versa for positive outcomes (Weinstein, 1980, 1983). Similarly, people tend to believe that future outcomes are more likely to be positive than contextual cues would suggest. For example, people are more likely to believe that other people are more likely to contract infectious diseases than themselves (e.g., Mccol et al., 2021; Salvador Casara et al., 2022).

Zero-risk bias

Zero-risk bias consists of the tendency to prefer the total elimination of a specific risk whereas the alternative option drives a greater decrease in overall risk (Schneider et al., 2017). For example, a study conducted by Viscusi and colleagues (1987) revealed that individuals tend to assign "a certainty premium" in order to totally eliminate a risk. Specifically, participants were asked to state the amount they would be willing to pay to reduce the risk of side effects from cleaning products, such as insecticide and toilet bowl cleaner. The results showed that people were willing to pay up to three times more for a reduction in risk from 5 out of 15,000 cases to 0 out of 15,000 cases compared to a risk reduction from 15 out of 15,000 cases to 10 out of 15,000 cases.

Logical fallacies

Argument from ignorance

This fallacy asserts that a proposition is true or false because there is no evidence to the contrary (Walton, 1999). For example, an individual can believe that vaccines cause autism because there is no specific evidence about vaccines not causing autism. A famous example that highlights the inconsistency of this argument is Russell's Teapot: if we cannot demonstrate that there is not a teapot orbiting around the Sun this does not mean that this teapot exists.

Argument to moderation

When somebody tends to believe that the truth stands in the compromise between opposite positions s/he falls for the argument to moderation fallacy (Halpern & Dunn, 2022). Indeed, there are no reasons for believing a priori that the moderation between different positions will lead to a more accurate view of an issue. If someone states that slavery is always wrong and someone else states that slavery is always right, this does not mean that slavery sometimes is wrong and sometimes right.

Dunning–Kruger Effect
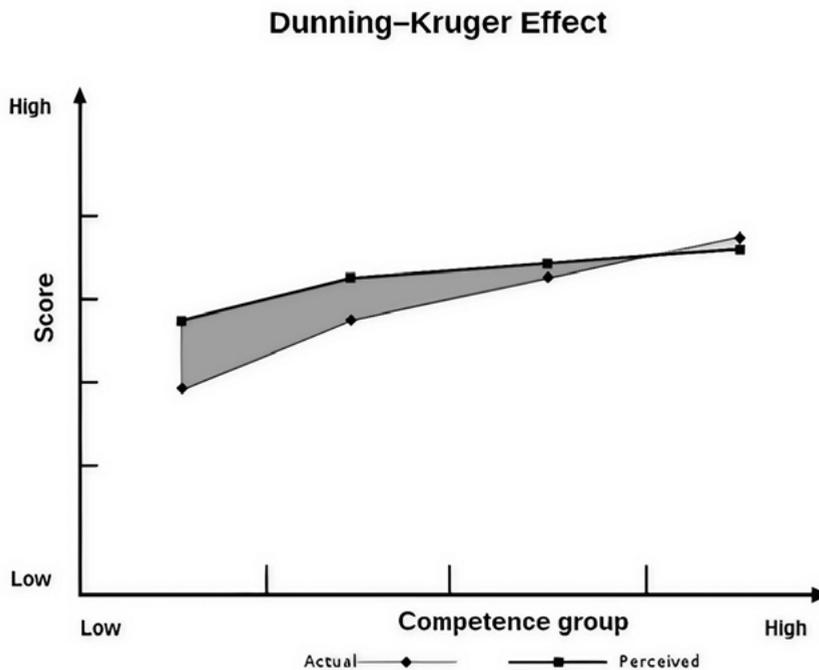
Figure 2. Graphical representation of the Dunning-Kruger effect. Low-competence individuals tend to overestimate their competence, whereas the opposite occurs for high-competence individuals.

Cum hoc ergo propter hoc

There is a correlation when changes in a variable are associated with changes in another variable. For example, a correlation may occur when:

1. When there is an increase/decrease of factor A, there is also an increase/decrease of factor B and viceversa.
2. When there is an increase/decrease of A, there is also a decrease/increase of B and viceversa

The cum hoc ergo propter hoc fallacy (Woods & Walton, 1977) is about believing that a mere observed association between A and B is due to A causing B. Actually, even if A and B are correlated there are several potential alternative explanations to the A causes B hypothesis:

1. It is B that causes A. The direction of the A causes B hypothesis to be wrong. For example, the sound of a locomotive steam whistle is not the cause of the functioning of the locomotive's engine but it is exactly the opposite.
2. A third factor C causes both A and B. For example, the increment of scientific progress may cause both an increase in available vaccines and an improvement in diagnostic disease tools. It may appear that with more available vaccines there are also more diseases, but in reality, there are only more diagnosed diseases thanks to better techniques.
3. The association between A and B is a mere coincidence. If we examined random factors the probability of finding correlations increases. Assuming that a correlation is due to a causal relation is, therefore, a rushed and potentially wrong conclusion.

False authority

False authority fallacy (Walton, 2010) results in appealing to a perceived expert in an unrelated field. For example, people may consider valid the arguments of a Nobel prize winner about political issues, even if he won the prize for discoveries in physics. In the health field, people appeal to false authority when they trust in discredited "experts" as homeopathic and alternative medicine practitioners.

Naturalistic fallacy

The outcome of the naturalistic fallacy is believing that natural things are healthy, morally right, and desirable, whereas artificial ones are noxious, immoral, and undesirable. In health-related contexts, chemicals, vaccines, and other medical procedures may be viewed as dangerous just because they are "unnatural" whereas herbs and alternative medicine procedures are natural. Actually, even natural remedies have side effects (Niggemann & Grüber, 2003; Vitalone, et al., 2011) and several artificial procedures improved Global Health (e.g., vaccination, techniques for purifying water).

Conclusions

Exploring the bounds of human cognition is a fascinating endeavor that has the potential to shape societies that are mindful of our short-comings and necessities. It leads us down a path of comprehending human nature in a more intricate way, as scientific research and empirical data demonstrate how cognitive resources confine us as social actors who search for coherence, meaning, and a sense of belonging. Such understanding has important implications related to how to shape the world through policies and to how environmental features emerge from our characteristics.

From a policy-making perspective, overcoming the limits of the expected utility hypothesis opens new modes of governance (Bell et al., 2010). Indeed, while policy-makers can tap into the utilitarian-oriented features of men and women by promoting behaviors targeting the incentive structures, for example, while governments can act to reduce the consumption of a specific good by increasing its price, they can also apply persuasion-based modes of governance. Specifically, some governments endorsed the idea of humans as cognitive misers in order to apply interventions (often called "nudges"; Thaler & Sunstein 2008) that exploit human heuristics and biases to promote behavioral change, like increase tax compliance (Jones et al., 2013). Following the same perspective, anchoring heuristic-based interventions are used to minimize tax resistance and to promote sustainable tourism (Kim et al., 2021). Moreover, cognitive-dissonance-based interventions are used to promote healthy behaviors such as the use of condoms (Stone et al., 1994). Finally, other widely used social interventions, for example, peer education (Green, 2001) takes into account the group dimension of humanity. Importantly, while biases can be used to promote behaviors leading to desirable outcomes, it is also possible that they can be used by marketing campaigns to fulfill goals that belong to private businesses. Think about the automatic renewals for Netflix or Amazon Prime subscriptions: these nudges are used not for the interest of the users but to maintain more subscriptions as possible.

While society can influence social behavior by exploiting cognitive biases, it is also important to notice that these psychological features influence and shape societies. One important example is related to online social media. The information that we gather on social media and the probability to engage with other people is not neutral, but it is built by specific algorithms programmed to capture our attention in the most effective way interacting with our psychological features (Aral, 2021). For

example, discussions and interactions among users tend to be clustered in echo chambers (e.g., Cinelli et al., 2021; Terren et al., 2021), which means that online spaces tend to cluster people with similar characteristics and similar attitudes. Such phenomenon is strongly based on the fact that people have ingroup biases and they are seekers of coherence, at the same time corporations managing social media base their business model on capturing people's attention; not surprisingly one good way to maintain people on the social media is to facilitate the interactions that users favor.

In conclusion, developing an understanding of human biases is critical for various reasons. It allows us to appreciate our limitations and explore ways to transcend them, defend ourselves from manipulation, and interpret how our cognition interacts with the world around us. By being aware of our cognitive limitations, we can create more effective policies, design better environmental outcomes, and ultimately build a more informed and equitable society.

References

Ajzen, I. (2011). The theory of planned behavior. *Organizational behavior and human decision processes*, *50*(2), 179-211. https://doi.org/10.1080/08870446.2011.613995

Aral, S. (2020). *The Hype Machine: How Social Media Disrupts Our Elections, Our Economy and Our Health – And How We Must Adapt*. New York: HarperCollins

Arkin, R. M., Appelman, A. J., & Burger, J. M. (1980). Social anxiety, self-presentation, and the self-serving bias in causal attribution. *Journal of Personality and Social Psychology, 38*(1), 23–35. https://doi.org/10.1037/0022-3514.38.1.23

Asch, S. E. (1956). Studies of independence and conformity: I. A minority of one against a unanimous majority. *Psychological Monographs: General and Applied, 70*(9), 1–70. https://doi.org/10.1037/h0093718

Baron, J. (1995). Myside bias in thinking about abortion. *Thinking & Reasoning, 1*(3), 221-235. https://doi.org/10.1080/13546789508256909

Cinelli, M., De Francisci Morales, G., Galeazzi, A., Quattrociocchi, W., & Starnini, M. (2021). The echo chamber effect on social media. *Proceedings of the National Academy of Sciences, 118*(9), e2023301118. https://doi.org/10.1073/pnas.2023301118

Cuddy, A. J., Fiske, S. T., & Glick, P. (2008). Warmth and competence

as universal dimensions of social perception: The stereotype content model and the BIAS map. *Advances in experimental social psychology, 40,* 61-149. https://doi.org/10.1016/S0065-2601(07)00002-0

Festinger, L. (1962). Cognitive dissonance. *Scientific American, 207*(4), 93-106.

Gilbert, D. T., & Malone, P. S. (1995). The correspondence bias. *Psychological bulletin, 117*(1), 21. https://doi.org/10.1037/0033-2909.117.1.21

Halpern, D. F., & Dunn, D. S. (2022). Critical Thinking. *Thought and Knowledge*, 1–32. https://doi.org/10.4324/9781003025412-1

Hannam, M., Hoy, C., Thompson, J.E. *et al.* General-relativistic precession in a black-hole binary. *Nature* (2022). https://doi.org/10.1038/s41586-022-05212-z

Hill, R. J., Fishbein, M., & Ajzen, I. (1977). Belief, Attitude, Intention and Behavior: An Introduction to Theory and Research. *Contemporary Sociology, 6*(2), 244. https://doi.org/10.2307/2065853

Jones, E. E., & Harris, V. A. (1967). The attribution of attitudes. *Journal of experimental social psychology, 3*(1), 1-24.

Kahneman, D., & Tversky, A. (2013). Prospect Theory: An Analysis of Decision Under Risk. *World Scientific Handbook in Financial Economics Series*, 99–127. https://doi.org/10.1142/9789814417358_0006

Kanouse, D. E., Hanson, L., Jones, E. E., Kelley, H., Nisbett, R., Valins, S., & Weiner, B. (1972). Attribution: Perceiving the causes of behavior. *Morristown, NJ: General Learning*, 47-62.

Kim, H. L., & Hyun, S. S. (2021). The anchoring effect of aviation green tax for sustainable tourism, based on the nudge theory. *Journal of Sustainable Tourism, 29*(7), 1082-1097. https://doi.org/10.1080/09669582.2020.1820017

Kruger, J., & Dunning, D. (1999). Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of Personality and Social Psychology, 77*(6), 1121–1134. https://doi.org/10.1037/0022-3514.77.6.1121

Lassiter, G. D., Geers, A. L., Munhall, P. J., Ploutz-Snyder, R. J., & Breitenbecher, D. L. (2002). Illusory causation: Why it occurs. *Psychological Science, 13*(4), 299-305. https://doi.org/10.1111/j.0956-7976.2002..x

Mccoll, K., Debin, M., Souty, C., Guerrisi, C., Turbelin, C., Falchi, A., ... & Raude, J. (2021). Are people optimistically biased about the risk of COVID-19 infection? Lessons from the first wave of the pandemic in Europe. *International Journal of Environmental Research and Public Health, 19*(1), 436. https://doi.org/10.3390/ijerph19010436

Miller, J. G. (1984). Culture and the development of everyday social explanation. *Journal of Personality and Social Psychology, 46*(5), 961–978. https://doi.org/10.1037/0022-3514.46.5.961

Nickerson, R. S. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of general psychology, 2*(2), 175-220. https://doi.org/10.1037/1089-2680.2.2.175

Niggemann, B., & Grüber, C. (2003). Side-effects of complementary and alternative medicine. *Allergy, 58*(8), 707-716. https://doi.org/10.1034/j.1398-9995.2003.00219.x

Ross, L., Lepper, M. R., & Hubbard, M. (1975). Perseverance in self-perception and social perception: Biased attributional processes in the debriefing paradigm. *Journal of Personality and Social Psychology, 32*(5), 880–892. https://doi.org/10.1037/0022-3514.32.5.880

Salvador Casara, B. G., Suitner, C., & Bettinsoli, M. L. (2019). Viral suspicions: Vaccine hesitancy in the Web 2.0. *Journal of Experimental Psychology: Applied, 25*(3), 354. https://doi.org/10.1037/xap0000211

Salvador Casara, B. G., Martinez-Conde, S., Dolinski, D., Suitner, C., Genschow, O., Muniak, P., & Kulesza, W. (2022). Trust in scientists, risk perception, conspiratorial beliefs, and unrealistic optimism: A network approach to investigating the psychological underpinnings of COVID-19 vaccination intentions. *Social Psychological Bulletin, 17*, 1-22. https://doi.org/10.32872/spb.7807

Schneider, E., Streicher, B., Lermer, E., Sachs, R., & Frey, D. (2017). Measuring the zero-risk bias: Methodological artefact or decision-making strategy? *Zeitschrift für Psychologie, 225*(1), 31–44. https://doi.org/10.1027/2151-2604/a000284

Schwarz, N., & Vaughn, L. A. (2002). The Availability Heuristic Revisited: Ease of Recall and Content of Recall as Distinct Sources of Information. *Heuristics and Biases*, 103–119. https://doi.org/10.1017/cbo9780511808098.007

Simon, H. A. (1957). *Models of man; social and rational.* Wiley

Slovic, P., & Lichtenstein, S. (1971). Comparison of Bayesian and regression approaches to the study of information processing in judgment. *Organizational behavior and human performance, 6*(6), 649-744. https://doi.org/10.1016/0030-5073(71)90033-X

Slovic, P., Finucane, M. L., Peters, E., & MacGregor, D. G. (2007). The affect heuristic. *European journal of operational research, 177*(3), 1333-1352. https://doi.org/10.1016/j.ejor.2005.04.006

Stillman, T. F., & Baumeister, R. F. (2009). Uncertainty, belongingness, and four needs for meaning. *Psychological Inquiry, 20*(4), 249-251. https://

doi.org/10.1080/1047840090333544

Stone, J., Aronson, E., Crain, A. L., Winslow, M. P., & Fried, C. B. (1994). Inducing hypocrisy as a means of encouraging young adults to use condoms. *Personality and Social Psychology Bulletin*, *20*(1), 116-128. https://doi.org/10.1177/0146167294201012

Tajfel, H. (1970). Experiments in Intergroup Discrimination. *Scientific American, 223*(5), 96–102. https://doi.org/10.1038/scientificamerican1170-96

Tajfel, H., Turner, J. C., Austin, W. G., & Worchel, S. (1979). An integrative theory of intergroup conflict. *Organizational identity: A reader*, *56*(65), 9780203505984-16.

Terren, L., & Borge-Bravo, R. (2021). Echo Chambers on Social Media: A Systematic Review of the Literature. *Review of Communication Research*, *9*, 99-118. Retrieved from https://rcommunicationr.org/index.php/rcr/article/view/94

Tversky, A., & Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. *Cognitive psychology*, *5*(2), 207-232. https://doi.org/10.1016/0010-0285(73)90033-9

Tversky, A., & Kahneman, D. (1974). Judgment under Uncertainty: Heuristics and Biases: Biases in judgments reveal some heuristics of thinking under uncertainty. *science*, *185*(4157), 1124-1131. DOI: 10.1126/science.185.4157.1124

Tversky, A., & Kahneman, D. (1983). Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment. *Psychological review*, *90*(4), 293. https://doi.org/10.1037/0033-295X.90.4.293

Van Zomeren, M. (2013). Four core social-psychological motivations to undertake collective action. *Social and Personality Psychology Compass*, *7*(6), 378-388. https://doi.org/10.1111/spc3.12031

Viscusi, W. K., Magat, W. A., & Huber, J. (1987). An investigation of the rationality of consumer valuations of multiple health risks. *The RAND journal of economics*, 465-479. https://doi.org/10.2307/2555636

Vitalone, A., Menniti-Ippolito, F., Moro, P. A., Firenzuoli, F., Raschetti, R., & Mazzanti, G. (2011). Suspected adverse reactions associated with herbal products used for weight loss: a case series reported to the Italian National Institute of Health. *European journal of clinical pharmacology*, *67*(3), 215-224. https://doi.org/10.1007/s00228-010-0981-4

Walton, D. (1999). The appeal to ignorance, or argumentum ad ignorantiam. *Argumentation*, *13*, 367-377.

Walton, D. (2010). *Appeal to expert opinion: Arguments from authority*. Penn State Press.

Weinstein, N. D. (1980). Unrealistic optimism about future life events. *Journal of personality and social psychology*, *39*(5), 806. http://dx.doi.org/10.1037/0022-3514.39.5.806

Weinstein, N. D. (1983). Reducing unrealistic optimism about illness susceptibility. *Health psychology*, *2*(1), 11. http://dx.doi.org/10.1037/0278-6133.2.1.11

Woods, J., & Walton, D. (1977). Post hoc, ergo propter hoc. *The Review of Metaphysics*, 569-593.

16. Auditory Gaydar: From Social Categorisation to Stigmatisation

Fabio Fasoli
University of Surrey, UK

Imagine you are at a bar with your friends, and you listen to a stranger talking. The man's voice makes you and your friend wonder whether he is gay. In doing so, you engage in the so-called 'auditory gaydar'. Gaydar is defined as the 'ability' to detect others' sexual orientation from minimal cues, including voice (Rule, 2017). As such, it involves a social categorisation process in which an individual is categorised as gay or heterosexual.[1] Often, gaydar judgments do not pass unnoticed and can have negative consequences on those who are the targets. This chapter aims to discuss literature on auditory gaydar and its consequences.

Auditory Gaydar

Sexual orientation is described as an 'ambiguous' social category that cannot be ascertained until the person self-discloses (see Tskhay & Rule, 2014). Still, people guess others' sexual orientation. Seminal work by Gaudio (1994) suggests that listeners judge male speakers' sexual orientation with some degree of accuracy. This finding prompted a series of other studies looking at a) whether auditory gaydar is accurate or not, b) whether voices of gay and straight speakers differ, and c) whether acoustic/phonetic cues associated with perceived sexual orientation can be identified.

---

1 Although different sexual orientations exist, the majority of gaydar literature has only focus on perception of individuals as gay or straight. This bias is reflected in the studies mentioned in this chapter.

Mixed results concerning accuracy have been found. Some studies suggest an overall, but far from perfect, auditory gaydar accuracy as speakers' sexual orientation was correctly identified from voice above a chance level (Linville, 1998; Pierrehumbert et al., 2004; Rieger et al., 2010; Valentova & Havlıcek, 2013). This was the case even when very little acoustic information (e.g., single consonant) was provided (Tracy et al., 2015). However, other studies have found an overall inaccuracy in auditory gaydar judgments (Munson et al., 2006; Painter et al., 2021; Smyth et al., 2003; Sulpizio et al., 2015, 2020). The results on acoustic cues have been similarly inconclusive. While some studies suggested that differences between gay and straight speakers' voices exist (Crist, 1997; Van Borsel et al., 2009), others highlighted the countless variations among gay and lesbian speakers' voices (Kachel et al., 2017, 2018). While some acoustic cues (e.g., sibilant /s/ in men, pitch in women) have been found to be associated with the perception of gay sexual orientation, it has become clear that some of those associations were influenced by voice-related stereotypes (e.g., lisping; see Mack & Munson, 2012). Nevertheless, literature shows that auditory gaydar judgments are not random. Indeed, listeners distinguish between speakers who are perceived to sound 'gay' and those who sound 'straight', regardless of their actual sexual orientation (see Fasoli et al., 2022; Sulpizio et al., 2015). Such judgments are guided by the so-called 'straight categorisation bias', suggesting that individuals assume everyone to be straight unless there are cues that suggest otherwise (Lick & Johnson, 2016). Individuals are quick and confident when they categorise speakers as heterosexual, but they are more hesitant to categorise them as gay (Sulpizio et al., 2015). Lick and Johnson (2016)indicated that cues of gender nonconformity make people move away from the 'straight' category. Gender nonconformity has indeed been found to be associated with auditory gaydar judgments, at least where heterosexual listeners are concerned (see Fasoli et al., 2022; Masi & Fasoli, 2022; Munson et al., 2006). As a consequence, a man who sounds feminine, with a high-pitched voice and a woman who has a low-pitched, more monotone and masculine-sounding voice, are more likely to be categorised as gay or lesbian, respectively.

The inconsistency of the results produced by auditory gaydar research needs to be considered in relation to methodological and theoretical issues. First, audio stimuli widely differed across studies as not only the number of speakers but also the nature and length of the stimuli varied (see Painter et al., 2021, Table 1). Speakers may sound different when

they utter a single vowel, read a sentence or text, or speak spontaneously. Also, shorter audio stimuli may prompt quick judgments, whereas longer audio recordings provide richer information that takes longer to be processed (see Painter et al., 2021).

Second, auditory gaydar judgments depend on how sexual orientation is conceptualised and assessed. Sexual orientation has often been measured as a binary concept by asking listeners to categorise speakers as either gay or straight (e.g., Linville, 1998; Smyth et al., 2003; Kachel et al., 2020). In such a case, an above chance level of correct responses was taken as an indicator of accuracy. As an alternative, a Kinsey-like scale asking participants to rate the speakers' sexual orientation from 1 (exclusively heterosexual) to 7 (exclusively gay/lesbian) has been used. This measure allows researchers to assess whether listeners differentiate speakers in relative terms (heterosexual and gay/lesbian speakers are judged differently) or in absolute terms (speakers are perceived differently but they are also placed on the correct side of the Kinsey like scale) (e.g., gay speakers are rated above 4; see Fasoli, Maass, & Berghella, 2022). Relative differences on a Kinsey-like scale have sometimes been interpreted as accurate gaydar (e.g., Valentova & Havlıcek, 2013). However, recent work has shown that relative differences on a Kinsey-like scale did not translate into accuracy on a binary choice (see Painter et al., 2021).

Third, auditory gaydar conceptualisation is based on the assumption that gaydar cues are 'static' (see Vasilovsky, 2018). However, voice is never static. It changes over time and across contexts thus affecting listeners' inferences of sexual orientation. Individuals can modulate their voices (see Crist, 1997). For instance, gay speakers can engage in a stereotypical 'gay speech' when they are among friends (see Podesva, 2007) but avoid it when they dislike such a stereotypical way of speaking (see Mann, 2012). Taking this into consideration, Daniele, Fasoli, Antonio, Sulpizio and Maass (2020) examined whether speakers modulated their voices depending on their 'coming out'. They found that gay speakers were perceived as more gay sounding when interacting with individuals they came out to and accepted them (e.g., best friend) than when talking with someone to whom they did not disclose their sexual orientation (e.g., grandparents). Moreover, they provided evidence that gay speakers were perceived as slightly more gay sounding after they had come out publicly on YouTube. These findings indicate that people can modulate their voices according to their identity and the situations they are in, thus questioning both the concept of gaydar as 'static' and its accuracy. Along with this, it needs

to be acknowledged that researchers often assumed that auditory gaydar functions similarly for male and female speakers. However, individuals endorse the idea that it is easier to detect sexual orientation for male speakers (Fasoli et al., 2018) and that vocal cues associated with a gay/lesbian sexual orientation are different (see Munson et al., 2006). This suggests that the processes may be different to some extent, but a very limited number of studies has compared auditory gaydar for the two genders (see Fasoli et al., 2022; Painter et al., 2021).

Sounding Gay/Lesbian and its consequences

The consequences of auditory gaydar occur in different forms, varying from stereotyping to discrimination. Most of the studies focused on the understanding of when and how the stigmatisation of gay-sounding speakers occurred. Still, little evidence exists on sexual minority individuals' experiences and expectations.

When someone listens to a speaker for a few seconds, they immediately form an impression. Such an impression is often based on stereotypes. For instance, gay-sounding male speakers are perceived as being stuck-up and outgoing (Tracy, 2019), sensitive and insecure, more likely to study 'feminine' subjects (e.g., psychology) and have 'feminine' hobbies (e.g., volleyball) or professions (e.g., hairdresser) (Fasoli, Maass, Paladino, & Sulpizio, 2017; Fasoli & Maass, 2018). On the other hand, lesbian-sounding female speakers are judged to engage in 'masculine' hobbies (e.g., football) and fields of study (e.g., engineering) more often than straight-sounding female speakers (Fasoli et al., 2017). Interestingly, these first impressions also affect the interpretation of what gay-sounding speakers say. For instance, when a gay-sounding speaker mentioned that he had been doing sports for many years, heterosexual listeners were more likely to imagine these sports to be ballet rather than football (Fasoli, Maass, Karniol, Antonio, & Sulpizio, 2020). Stereotypical attributions go even further and make people guess the speakers' health status. Fasoli, Maass, and Sulpizio (2018) asked heterosexual participants to listen to gay/lesbian- and straight-sounding speakers and rate their likelihood of suffering from diseases stereotypically associated with men (e.g., alcoholism) or women (e.g., anorexia) along with stereotypical diseases associated with sexual minorities (e.g., HIV). They found that, compared to straight-sounding speakers of the same gender, gay-sounding men were believed to be more likely to suffer from diseases associated with wom-

en and sexual minorities, while lesbian-sounding women were judged to suffer more often from men's diseases. Overall, this line of research shows that a voice triggers a variety of stereotypical attributions to gay/lesbian-sounding speakers.

Beyond stereotyping, speakers perceived to be gay are more likely to be discriminated against. For instance, students who identified as heterosexual but sounded gay were stigmatised when applying for a scholarship, possibly because their voices violated the expectations that come along with identifying as straight (Gowen & Britt, 2006). Not only students, but also teachers who sound gay, face consequences. Recent work has shown that gay-sounding teachers are perceived as less effective in their teaching than straight-sounding ones (Taylor & Raadt, 2021). Expanding these findings to other contexts, research has focused on voice-based discrimination during the hiring process. For instance, gay-sounding speakers applying for top jobs (e.g., CEO) were perceived as less gender conforming and, thus, as less suitable for the position than straight-sounding speakers (Fasoli et al., 2017). Interestingly, this effect was found when heterosexual participants were exposed to the applicants' voices, but not when they were exposed to their (gay/straight-looking) faces, suggesting that voice may be more powerful than facial cues in such a context (see Rakic et al., 2011 for similar results concerning other social categories). This voice-discrimination bias was replicated across different countries (Italy, the United Kingdom, Brazil) and seems to be particularly strong among highly prejudiced heterosexual individuals (see Fontenele, Souza, & Fasoli, 2022). However, less is known about discrimination differences that may occur between gay- and lesbian-sounding speakers in the hiring process. As mentioned earlier, research suggests that individuals generally believe it is easier to detect the sexual orientation of men than of women (Fasoli et al., 2018), which implies that listeners may be more likely to pick up on gay-sounding men and discriminate against them more. However, research on voice-discrimination in hiring decisions comparing gay-sounding and lesbian-sounding applicants (Fasoli & Hegarty, 2020) has shown that lesbian-sounding women bore the brunt of discrimination. They were perceived as the ones most likely to lack competence and thus were not seen to be suitable for managerial roles. This result was interesting for different reasons. First, it showed that voice signalling a (gender and sexual orientation) double minority could negatively affect career opportunities. Second, it suggests that lesbian-sounding women are at a higher risk of discrimination even though people believe auditory

gaydar is more a thing for men more than for women (Fasoli, Hegarty, & Frost, 2021). Third, auditory gaydar seems to be related to power issues. Indeed, in this research it was also found that gay- and lesbian-sounding speakers were perceived to be better suited for low-status positions (e.g., manager's assistant) than straight-sounding speakers. This implies that when a voice signals a minority status, heterosexual listeners may prevent those speakers to reach a higher status position. This is important considering the fact that voice-based discrimination is 'subtle' and difficult to prove (see Castle, 2009).

Another interesting context in which stereotyping and discrimination came to light had to do with the perception of gay- and straight-sounding speakers as potential adoptive parents (Fasoli & Maass, 2020). Here, gay-sounding male speakers were rated as warmer and having better parental skills than straight-sounding male speakers, possibly because listeners relied on the stereotypical idea that gay men hold more 'feminine' characteristics. However, when heterosexual participants had to decide who should be the adoptive parent, in some cases, they preferred the straight-sounding over the gay-sounding speaker, showing a discriminatory bias. Indeed, while the impressions of speakers were similar for heterosexual participants in Italy and the United Kingdom, the discriminatory bias was only found among Italian participants. Considering that, differently to the United Kingdom, same-sex adoption is illegal in Italy, this effect may indicate that the normative context in which decisions are made matters, and that equality-based laws can prevent or even decrease voice-based biases.

Finally, it is important to look at the speakers' perspectives. Unfortunately, the literature in this regard is limited. It has been found that sexual minority men and women, who believe they sound less gender-conforming (less masculine for men and less feminine for women) are aware that they are more likely to be perceived as gay or lesbian because their voices match gender-inverted stereotypes. Moreover, sexual minority men do not desire for their voices to disclose their sexual orientation to strangers as this may lead to stigmatisation (see Fasoli, Hegarty, Maass, & Antonio, 2018). Indeed, lesbian women, and specifically gay men, who believe they sound gay/lesbian expect to be stigmatised because of the way they sound. They are also vigilant about the people around them and the comments they may make about their voices (Fasoli, Hegarty, & Frost, 2021). Although more research is needed, these findings suggest that sounding gay can be difficult and may cause stress to individuals as they expect to be stigmatised.

Conclusion

Research on auditory gaydar has shown that this voice-based social categorisation process is complex. It relies on assumptions about heteronormativity (i.e., heterosexuality as the norm) and 'stability' and holds on to voice-based stereotypes concerning gender and sexual orientation. Also, auditory gaydar cannot be seen as a 'neutral' phenomenon in which listeners simply guess others' sexual orientation, since it triggers stereotyping and discrimination against gay/lesbian-sounding speakers.

References

Alt, N. P., Lick, D. J., & Johnson, K. L. (2020). The straight categorization bias: A motivated and altruistic reasoning account. *Journal of Personality and Social Psychology*, *119*(6), 1266-1289.

Castle, R. (2012). The gay accent, gender, and title VII employment discrimination. *Seattle University Law Review, 36*, 1943–1966.

Crist, S. (1997). Duration of onset consonants in gay male stereotyped speech. *University of Pennsylvania Working Papers in Linguistics*, *4*(3), 53-70.

Daniele, M., Fasoli, F., Antonio, R., Sulpizio, S., & Maass, A. (2020). Gay voice: Stable marker of sexual orientation or flexible communication device?. *Archives of Sexual Behavior*, *49*(7), 2585-2600.

Fasoli, F., & Hegarty, P. (2020). A leader doesn't sound lesbian!: The impact of sexual orientation vocal cues on heterosexual persons' first impression and hiring decision. *Psychology of Women Quarterly*, *44*(2), 234-255.

Fasoli, F., Hegarty, P., & Frost, D. M. (2021). Stigmatization of 'gay-sounding' voices: The role of heterosexual, lesbian, and gay individuals' essentialist beliefs. *British Journal of Social Psychology*, *60*(3), 826-850.

Fasoli, F., Hegarty, P., Maass, A., & Antonio, R. (2018). Who wants to sound straight? Sexual majority and minority stereotypes, beliefs, and desires about auditory gaydar. *Personality and Individual Differences*, *130*, 59-64.

Fasoli, F., & Maass, A. (2018). Voice and prejudice: The social costs of auditory gaydar. *Atlantic Journal of Communication*, *26*(2), 98-110.

Fasoli, F., & Maass, A. (2020). The social costs of sounding gay: Voice-based impressions of adoption applicants. *Journal of Language and Social Psychology*, *39*(1), 112-131.

Fasoli, F., Maass, A., & Berghella, L. (2022). Who has a better auditory

gaydar? Sexual orientation categorization by heterosexual and lesbian, gay and bisexual people. *Journal of Homosexuality*. Advanced online publication: https://doi.org/10.1080/00918369.2021.2004796

Fasoli, F., Maass, A., Paladino, M. P., & Sulpizio, S. (2017). Gay-and lesbian-sounding auditory cues elicit stereotyping and discrimination. *Archives of Sexual Behavior*, *46*(5), 1261-1277.

Fasoli, F., Maass, A., & Sulpizio, S. (2018). Stereotypical disease inferences from gay/lesbian versus heterosexual voices. *Journal of Homosexuality*, *65*(8), 990-1014.

Fontenele, A. B. G., Souza, L. E. C. D., & Fasoli, F. (2022). Who does discriminate against gay-sounding speakers? The role of prejudice on voice-based hiring decisions in Brazil. *Journal of Language and Social Psychology*. Advanced online publication: https://doi.org/10.1177%2F0261927X221077243

Gaudio, R. P. (1994). Sounding gay: Pitch properties in the speech of gay and straight men. *American Speech*, *69*(1), 30-57.

Gowen, C. W., & Britt, T. W. (2006). The interactive effects of homosexual speech and sexual orientation on the stigmatization of men: Evidence for expectancy violation theory. *Journal of Language and Social Psychology*, *25*(4), 437-456.

Kachel, S., Simpson, A. P., & Steffens, M. C. (2017). Acoustic correlates of sexual orientation and gender-role self-concept in women's speech. *The Journal of the Acoustical Society of America*, *141*(6), 4793-4809.

Kachel, S., Radtke, A., Skuk, V. G., Zäske, R., Simpson, A. P., & Steffens, M. C. (2018). Investigating the common set of acoustic parameters in sexual orientation groups: A voice averaging approach. *PloS one*, *13*(12), e0208686

Kachel, S., Steffens, M. C., Preuß, S., & Simpson, A. P. (2020). Gender (conformity) matters: Cross-dimensional and cross-modal associations in sexual orientation perception. *Journal of Language and Social Psychology*, *39*(1), 40-66.

Lick, D. J., & Johnson, K. L. (2016). Straight until proven gay: A systematic bias toward straight categorizations in sexual orientation judgments. *Journal of Personality and Social Psychology*, *110*(6), 801-817.

Linville, S. E. (1998). Acoustic correlates of perceived versus actual sexual orientation in men's speech. *Folia Phoniatrica et Logopaedica*, *50*(1), 35-48.

Mack, S., & Munson, B. (2012). The influence of/s/quality on ratings of

men's sexual orientation: Explicit and implicit measures of the 'gay lisp' stereotype. *Journal of Phonetics*, *40*(1), 198-212.

Mann, S. L. (2012). Speaker attitude as a predictive factor in listener perception of gay men's speech. *Journal of Language and Sexuality*, *1*(2), 206-230.

Masi, M., & Fasoli, F. (2022). When fluency matters: The interplay between categorization fluency and gender atypicality on gaydar judgments. *Journal of Language and Social Psychology*. Advanced online publication: https://doi.org/10.1177%2F0261927X221111382

Munson, B., McDonald, E. C., DeBoe, N. L., & White, A. R. (2006). The acoustic and perceptual bases of judgments of women and men's sexual orientation from read speech. *Journal of Phonetics*, *34*(2), 202-240.

Painter, D., Fasoli, F., & Sulpizio, S. (2021). The impact of stimuli length and analytic method on auditory 'gaydar' research. *Journal of Voice*. Advanced online publication: https://doi.org/10.1016/j.jvoice.2021.08.016

Pierrehumbert, J. B., Bent, T., Munson, B., Bradlow, A. R., & Bailey, J. M. (2004). The influence of sexual orientation on vowel production (L). *The Journal of the Acoustical Society of America*, *116*(4), 1905-1908.

Podesva, R. J. (2007). Phonation type as a stylistic variable: The use of falsetto in constructing a persona 1. *Journal of Sociolinguistics*, *11*(4), 478-504.

Rakić, T., Steffens, M. C., & Mummendey, A. (2011). Blinded by the accent! The minor role of looks in ethnic categorization. *Journal of Personality and Social Psychology*, *100*(1), 16-29.

Rieger, G., Linsenmeier, J. A., Gygax, L., Garcia, S., & Bailey, J. M. (2010). Dissecting "gaydar": Accuracy and the role of masculinity–femininity. *Archives of Sexual Behavior*, *39*(1), 124-140.

Rule, N. O. (2017). Perceptions of sexual orientation from minimal cues. *Archives of Sexual Behavior*, *46*(1), 129-139.

Smyth, R., Jacobs, G., & Rogers, H. (2003). Male voices and perceived sexual orientation: An experimental and theoretical approach. *Language in Society*, *32*(3), 329-350.

Sulpizio, S., Fasoli, F., Maass, A., Paladino, M. P., Vespignani, F., Eyssel, F., & Bentler, D. (2015). The sound of voice: Voice-based categorization of speakers' sexual orientation within and across languages. *PloS one*, *10*(7), e0128882.

Sulpizio, S., Fasoli, F., Antonio, R., Eyssel, F., Paladino, M. P., & Diehl, C. (2020). Auditory gaydar: Perception of sexual orientation based on

female voice. *Language and Speech*, *63*(1), 184-206.

Taylor, D. M., & Raadt, J. S. (2021). Gay-and straight-sounding auditory cues elicit stereotyping about teaching effectiveness. *Journal of Research in Music Education*, *69*(1), 62-84.

Tracy, E. C. (2019). Judgments of American male talkers who are perceived to sound gay or heterosexual: Certain social contexts don't affect perception of personality traits. *The Journal of the Acoustical Society of America*, *146*(4), 3056-3056.

Tracy, E. C., Bainter, S. A., & Satariano, N. P. (2015). Judgments of self-identified gay and heterosexual male speakers: Which phonemes are most salient in determining sexual orientation?. *Journal of Phonetics*, *52*, 13-25.

Tskhay, K. O., & Rule, N. O. (2013). Accuracy in categorizing perceptually ambiguous groups: A review and meta-analysis. *Personality and Social Psychology Review*, *17*(1), 72-86.

Valentova, J. V., & Havlíček, J. (2013). Perceived sexual orientation based on vocal and facial stimuli is linked to self-rated sexual orientation in Czech men. *PloS one*, *8*(12), e82417.

Van Borsel, J., Janssens, J., & De Bodt, M. (2009). Breathiness as a feminine voice characteristic: A perceptual approach. *Journal of Voice*, *23*(3), 291-294.

Vasilovsky, A. T. (2018). Aesthetic as genetic: The epistemological violence of gaydar research. *Theory & Psychology*, *28*(3), 298-318.

17. Implicit social cognition

Maddalena Marini
University of Campania "Luigi Vanvitelli", Italy

Introduction

As human beings, we like to believe that we have full control of our behavior. However, decades of research on mental functions have shown that this is not the case: many of our decisions and actions occur in an automatic fashion (Wheatley & Wegner, 2001).

At any given moment, our mind processes information without our intentional effort, influencing our behavior. For example, it can automatically determine the way we pay attention to input from the environment and process the world around us, as well as how we perceive and interact with other people (Bargh & Williams, 2006).

Research showed indeed that we can automatically form our impressions and judgments about others based solely on their physical, social, and psychological characteristics (e.g., facial configuration, skin color, gender, religion, sexuality, disability, and personality) (Nosek et al., 2002; Olivola et al., 2014). For instance, we tend to judge the face on the right in Figure 1 as more competent than the face on the left based on its facial features.

This discovery profoundly influenced the field of social psychology as it provided evidence that social psychological constructs, such as stereotypes, attitudes, and self-concepts, can operate automatically with relatively little control over behavior.

Implicit social cognition is the study of attitudes, stereotypes, and self-concepts that occur outside of intentional control (Greenwald & Ba-

naji, 1995). This discipline posed the basis for the development of an exponential growth of research characterized by a proliferation of novel measurement instruments captured by the term *implicit measures* (Age of Measurement; Nosek et al., 2011). Unlike explicit measures (i.e., self-reports), implicit measures aim to infer social psychological constructs indirectly, that is without instruction to report it, assuming no introspective of the construct (Greenwald et al., 2019; Greenwald & Banaji, 2017; Greenwald & Lai, 2020)

The present chapter aims to provide an overview of the most popular paradigms developed to measure implicit social cognition and their main findings in this research field. Specifically, we will discuss their relationship with explicit measures, their pervasiveness, their predictive validity in relation to behavior, their malleability, and their stability over time.
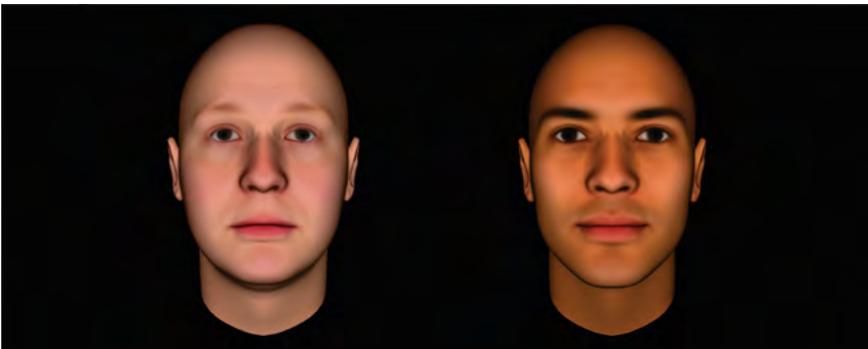
Figure 1. Facial appearance shapes social attributions and impressions. For instance, the face below on the right is perceived as more competent than the face on the left.
Note. Adapted from Olivola et al. (2014).

Paradigms

Implicit measures were developed to overcome the limitations of explicit measures (i.e., self-reports) in assessing metal contents (Greenwald & Banaji, 1995).

Explicit measures assess constructs of interest directly by asking people their thoughts and feelings. For example, in the Racial Attitudes Scale (RAS; Sidanius et al., 1991), attitudes towards racial and ethnic groups are assessed using a series of statements that participants rate to indicate their level of agreement or disagreement. These statements typically ad-

dress issues related to racial equality, stereotypes, racial discrimination, social distance, and preferences for certain racial or ethnic groups.

Although these instruments are very useful to address several psychological questions (Blanton & Jaccard, 2018), they are unable to detect mental contents that people may be incapable or unwilling to report.

Suppose, for instance, that you are interested in knowing if a person is a fan of the New York Yankees or the Boston Red Sox. People usually have no difficulties in reporting their preferences for one sports team over another. However, if you are interested in investigating sensitive social issues, such as if a person prefers a social group compared to another social group (e.g., preference for White people over Black people), people may be unable or reluctant to report their preferences.

Explicit measures, indeed, rely exclusively on introspective experience, i.e., the direct examination of one's own mental contents, and thus they are not suited to capture thoughts and feelings that are outside of conscious awareness (Greenwald & Banaji, 1995). In addition, they are sensitive to social desirability and self-representation processes (Crowne & Marlowe, 1960). In other words, people may not report their own preferences because they are unaware of them (i.e., the mental content may be inaccessible to introspection) or if they think this could be viewed negatively by others.

Unlike explicit measures, implicit measures assess mental content indirectly by means of performance-based instruments. That is, they use performance parameters in specific tasks (e.g., response latency and errors) to infer mental contents, limiting participants' ability to control their responses. Implicit measures are thus less influenced by social desirability and self-representation processes and do not rely on introspection for measurement of thoughts and feelings (Greenwald & Banaji, 1995).

Implicit measures were originally defined as tools to detect psychological constructs that occur outside of conscious awareness (Greenwald & Banaji, 1995). The term implicit thus referred to proprieties of mental processes or mental representations that may operate in an unconscious fashion. This definition of implicit derived from memory studies of the 1980s in which indirect measures were used to detect operations of memory that occurred without conscious recollection (Richardson-Klavehn & Bjork, 1988). However, subsequent research in this field (Jacoby, 1991; Reingold & Merikle, 1988) posed a challenge to the term "implicit" as unconscious, leading to the definition of implicit as proprieties of psychological measures that detect a construct indirectly (implicitly) versus

directly (explicitly). This research provided convincing arguments that was neither appropriate to define indirect measures as pure indicators of unconscious processes nor to define direct measures as pure indicators of conscious processes.

In line with this research, in this chapter, the term implicit is used to describe the measures used for the assessment of psychological constructs and not to define the nature of the constructs measured. I use the term "implicit" to describe the large class of indirect measures in social cognition: measures in which the construct is inferred from the respondents' performance at tasks without asking them for a direct report or introspection (Greenwald & Lai, 2020). Similarly, the adjective implicit referred to the psychological constructs described in this chapter (e.g., attitudes, stereotypes, associations) is used to indicate that those constructs are measured indirectly by means of implicit measures and not as synonymous of unconscious.

Over the past three decades, research in implicit social cognition produced a proliferation of paradigms and instruments described by the term implicit measures (Nosek et al., 2011). In the following, I describe the two main instruments (and their variants) that have been more frequently used and provided the largeness pool of research: the sequential priming (Fazio et al., 1995) and the Implicit Association Test (Greenwald et al., 1998). In addition, I illustrate additional instruments that, although less utilized, provided valuable insights into implicit social cognition research. These include the Implicit Relational Assessment Procedure (Barnes-Holmes et al., 2006), the Linguistic Intergroup Bias (Maass et al., 1989), and the Approach-avoidance paradigms (Chen & Bargh, 1999; Eder & Rothermund, 2008).

Sequential priming procedures

The first paradigm employed to assess indirectly mental contents was sequential priming (Fazio et al., 1986; Wentura & Degner, 2010), which evaluates the effect of a first stimulus (prime) on the processing of a second stimulus (target). In a classic priming task, participants are rapidly presented with a prime and asked to respond to a target that appears shortly after. The underlying assumption is that when the target and the prime are congruent with mental contents, responses are faster and more accurate than when they are incongruent. For example, a person who holds preferences for White people over Black people should show faster and more accurate responses in classifying the valence of positive words

when these are previously primed with a picture of a White person than a picture of a Black person (Fazio et al., 1995) (see Figure 2). Similarly, a person who holds gender stereotypes should show faster and more accurate responses in classifying female pronouns when these are previously primed with stereotypically female professions (e.g., nurse) than with stereotypically male professions (e.g., doctor) (Banaji & Hardin, 1996).

Several variants of sequential priming can be identified based on the required response to the target. These include the evaluative decision task, which requires to classify the target based on its positive or negative valence (Fazio et al., 1995), the lexical decision task, which entails distinguishing the target as a word or a nonword (Wittenbrink et al., 1997), and the semantic decision task, which involves the evaluation of the target in terms of its semantic proprieties (e.g., gender of pronouns) (Banaji & Hardin, 1996; Blair & Banaji, 1996).

A characteristic variant of sequential priming is the Affective Misattribution Procedure (AMP; Payne et al., 2005). Unlike the other priming tasks, in the AMP participants are asked to report their subjective evaluations (i.e., they like or dislike) in response to ambiguous target stimuli (e.g., Chinese pictographs for non-Chinese respondents). The basic idea is that the targets are judged more favorably when they are primed with a positive than a negative stimulus. That is, primes should misattribute participants' affective feelings to the targets. For example, participants with a preference for White people over Black people should show a tendency to evaluate Chinese pictographs as more pleasant when they are primed with a picture of a White individual than when they are primed with a picture of a Black individual (Payne & Lundberg, 2014).
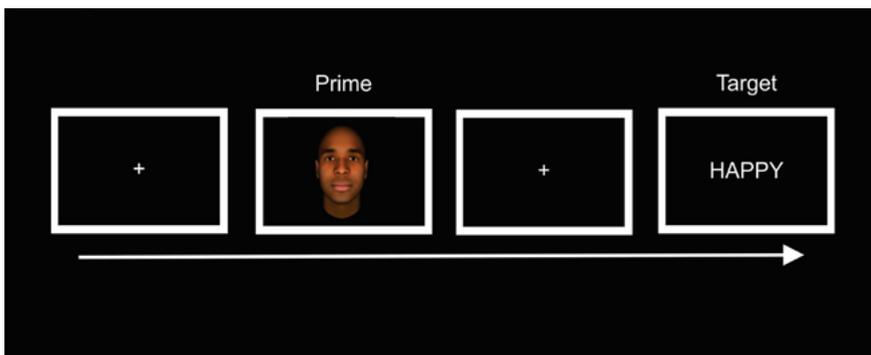

Figure 2. Illustration of a sequential priming procedure to assess racial attitudes.

Implicit Association Test and its variants

The IAT is the most widely used measure in implicit social cognition research (Epifania et al., 2022). It assesses mental contents by measuring how quickly and accurately a person can categorize and associate stimuli in two different sorting conditions. The underlying assumption is that stimuli that are strongly associated at a mental representation level show shorter reaction times and fewer errors when classified together than when they are not (Greenwald et al., 1998; Nosek, Greenwald, et al., 2007).

For example, in a typical IAT assessing racial attitudes (i.e., race IAT; Figure 3), participants categorize stimuli representing the two conceptual categories – White people and Black people – and the two evaluative attributes – good and bad. In one sorting condition (congruent condition with the racial attitude), participants categorize pictures of White people and good words (e.g., joy, love, and peace) using one response key, while pictures of Black people and bad words (e.g., agony, terrible, and horrible) are categorized with another response key. In the other sorting condition (incongruent condition with the racial attitude), pictures of White people and bad words are categorized with one key, whereas pictures of Black people and good words are categorized using the other response key. Shorter categorization times in the first sorting condition compared to the second sorting condition are indicative of a preference for White people over Black people. Conversely, longer categorization times in the second sorting condition compared to the first one are indicative of a preference for Black people over White people.

Variations of the IAT include the Brief-IAT (B-IAT; Marini et al., 2021, 2022; Sriram & Greenwald, 2009), the Single-Category IAT (or Single-Target IAT; (Bluemke & Fiedler, 2009; Karpinski & Steinman, 2006), the Go/No-Go Association Task (GNAT; Nosek & Banaji, 2001), and the Multi-Category IAT (MC-IAT; Marini, 2017; Nosek et al., 2014). The Brief-IAT (B-IAT) is a shortened version of the standard IAT that requires focusing only on one focal category and attribute in each sorting condition. For example, participants press one key for pictures of White people and good words and another key when "anything else" appears. The Single Category IAT (SC-IAT), also known as Single-Target IAT (ST-IAT), uses only one category (e.g., Black people) and two attributes (e.g., good and bad) to measure the mental content toward a single target object (e.g., attitudes towards Blacks). Similarly, the Go/No-go Association Test (GNAT) assesses the strength of evaluative associations with individual

target objects by asking participants to press a key (go) in response to some stimuli (e.g., pictures of Black people and negative words), and withhold a response to other stimuli (e.g., pictures of White people, positive words, and distractor stimuli). Finally, the Multi-Category IAT (MC-IAT) compares mental contents between more than two categories (e.g., Blacks, Whites, Asians, and Hispanics).
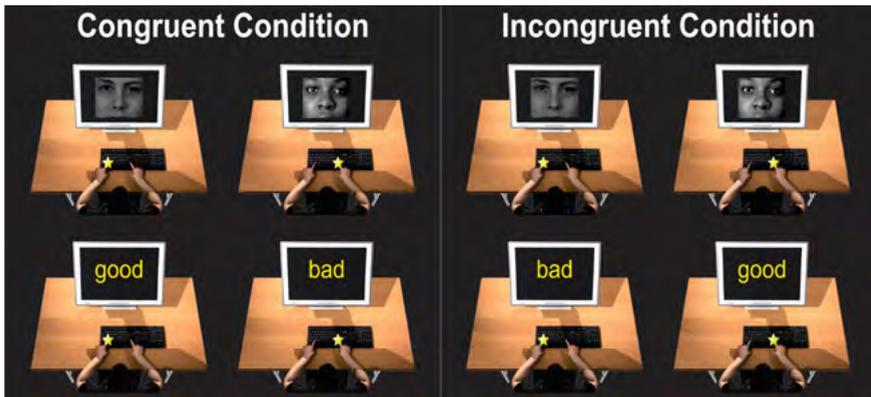
Figure 3. Illustration of a race Implicit Association Test (IAT).
Note. Adapted from Marini et al. (2018). Congruent and incongruent labels are used to define the conditions in which motor responses required in the task are, respectively, compatible or incompatible with the racial attitude.

Other paradigms

Implicit Relational Assessment Procedure

The Implicit Relational Assessment Procedure (IRAP; Barnes-Holmes et al., 2006) assesses mental contents by measuring the propositions between stimuli. For example, in an IRAP assessing weight attitudes, participants are presented with a picture of either a thin or an overweight individual along with either a positive or a negative attribute. They are required to select a relational response option such as "true" or "false" in two different conditions. In one condition, they are required to confirm the thin-positive relation, while in another, they are asked to confirm the overweight-positive relation. The basic hypothesis is that participants are faster when the relation required in the two conditions is consistent with their beliefs. That is, faster responses in the condition confirming

the thin-positive relation than the overweight-positive relation are interpreted as a preference for thin people over overweight people.

Linguistic Intergroup Bias

The Linguistic Intergroup Bias (Maass et al., 1989, 1995, 1996) is an indirect measure that evaluates self-reported responses to assess mental contents. The underlying idea is that people tend to use abstract language to describe actions that they believe to be stereotypical of a certain group (stereotypical-consistent events), and concrete language to describe unusual or uncharacteristic behavior (stereotypical-inconsistent events). For example, a positive action performed by an in-group member is more likely to be described in abstract terms because it is viewed as characteristic and typical, while a negative action is described in concrete terms because it is viewed as uncharacteristic. Conversely, a positive action performed by an out-group member is more likely to be described in concrete terms because it is viewed as uncharacteristic, while a negative action is described in abstract terms because it is viewed as characteristic and typical (see also Chapters XX, in this book).

Approach-avoidance tasks

These paradigms are based on the idea that positive stimuli elicit spontaneous approach reactions, while negative stimuli elicit spontaneous avoidance reactions (Chen & Bargh, 1999; Eder & Rothermund, 2008). For example, participants should be faster at pushing a lever towards them (approach) in response to positive stimuli and pushing it away from them (avoidance) for negative stimuli. Thus, a person who holds preferences for White people over Black people should be faster in making an approach movement in response to stimuli of in-group members and an avoidance movement in response to stimuli of out-group members (Paladino & Castelli, 2008).

Choosing a paradigm

Selection of an implicit measure requires the consideration of different aspects.

First, it is necessary to carefully select the implicit measure that most closely meet the specific construct you are trying to assess and the conditions that characterize it. This means considering the potential limita-

tions of each measure. For example, the IAT suffers from a constraint as it provides a combined measure of associations between two categories and two attributes that cannot be decomposed. That is, using an IAT, it is not possible to known whether the score is primarily driven by one of the two associations assessed. If you use an IAT to assess racial attitudes, for instance, you cannot say whether the score reflects a stronger pro-White preference (i.e., White people + good associations) or a negativity toward Black people (i.e., Black people + bad associations). All you can say with confidence is that there is a stronger association of White people with good and Black people with bad. Thus, if you are interested to assess an attitude about a single concept, it is more appropriate to use sequential priming tasks or variants of the standard IAT that have been created to address this issue (e.g., SC-IAT). Indeed, unlike the IAT, primes sequential priming tasks permit the calculation of separate scores for each of the four associations (Wentura & Degner, 2010). Similarly, the SC-IAT allows to assess constructs for single targets (Bluemke & Fiedler, 2009; Karpinski & Steinman, 2006).

Another relevant aspect to consider is related to the psychometric proprieties of the implicit measure. Research showed indeed that not all implicit measures have the same psychometric proprieties (Bar-Anan & Nosek, 2014; Greenwald & Lai, 2020). Specifically, it has been demonstrated that internal consistency varied greatly among instruments: most of them show acceptable or good reliability, such as the IAT and the AMP (e.g., IAT: $\alpha= 0.80$; AMP $\alpha= 0.82$), while others yield poor or questionable internal consistency (e.g., IRAP: $\alpha= 0.66$; evaluative decision task: $\alpha= 0.53$) (Greenwald & Lai, 2020). For example, sequential priming procedures have been criticized for their low reliability, which rarely exceed Cronbach's Alpha value of 0.50 (Gawronski & De Houwer, 2014), despite it is among the most widely used instruments in research using implicit measures. Similarly, only some of these measures showed test-retest reliability estimate around r=0.50 (AMP: r=0.56; IAT: r=0.49). For example, approach-avoidance tasks showed satisfactory estimates of internal consistency, but their reliability varies considerably depending on the variant that is used (Krieglmeyer & Deutsch, 2010).

Main findings

Relation with explicit measures

Although implicit and explicit measures tend to show positive correlations, they can display a large degree of dissociation or even opposite effects (Greenwald & Nosek, 2008; Marini, 2017; Marini et al., 2013; Nosek, 2007). For example, a study investigating the preferences toward different weight categories (i.e., underweight, normal weight, overweight/obese), showed that, when the underweight category was contrasted with the obese/overweight category, results differed at the explicit and implicit levels although scores were positively correlated. That is, pro-underweight preferences were observed at the explicit level, while pro-overweight/obese preferences were found at the implicit level (Marini, 2017). This finding suggests that explicit and implicit weight preferences may be influenced by different social and cultural factors. Explicit preferences may reflect standards of own culture and society as they are more influenced by intentional and social desirability processes. The pro-underweight preferences observed at the explicit level may be thus attributed to the social pressures and cultural ideal of thinness prevalent in our society. In contrast, implicit preferences may reflect basic needs such as that of self-preservation and survival of the individual firmly established in our mind as they infer automatic processes and mechanisms. Therefore, the pro-overweight/obese preferences found at the implicit level may stem from concerns about the health consequences of extreme low body weight or be a result of evolutionary mechanisms that detect and stigmatize markers of disease, with individuals who are very thin more likely to elicit disease-related cognitions and emotions (Marini, 2017).

Similarly, a study with more than 700,000 respondents demonstrated that while White individuals endorsed egalitarian views at the explicit level and pro-White attitudes at the implicit level, implicit and explicit measures showed moderate positive correlations (Nosek, Smyth, et al., 2007).

Taken together these findings indicate some shared and independent variance between implicit and explicit measures, suggesting thus that these instruments assess related but distinct constructs (Cunningham et al., 2004; Nosek, Smyth, et al., 2007).

Dissociations between implicit and explicit measures posed the basis for numerous theories about the nature of processes underlying these instruments.

Dual-process theories assumed that implicit and explicit measures reflect the outcomes of two qualitatively distinct processes: automatic associative processes for implicit measures and controlled reasoning processes for explicit measures (Fazio, 2007; Gawronski & Bodenhausen, 2006; Rydell & McConnell, 2006; Strack & Deutsch, 2004). For example, the Motivation and Opportunity as Determinants (MODE) model (Fazio, 2007) proposed that implicit measures capture the activation of automatic associations in response to an object, whereas explicit measures tap these retrieved associations along with other cognitive processes and mental contents that can be activated by the context or stem by people's motivation and opportunity in engaging in deliberative processes. Thus, dissociations or low correlations between implicit and explicit measures should occur because people engage in some deliberative processing that modifies the automatic associations initially activated. Similarly, the Associative-Propositional Evaluation (APE) model (Gawronski & Bodenhausen, 2006) assumed that implicit measures reflect activation of mental associations, while explicit measures reveal the outcome of propositional processes that assess the validity of these activated mental contents. If the associations captured by implicit measures are inconsistent with information derived by propositional reasoning, people may reject these associations to restore cognitive consistency, leading to dissociations or low correlations between implicit and explicit measures.

Over the past decade, dual-process theories have been challenged by single-process theories, which explain the outcomes of implicit and explicit measures as the product of a unique propositional process (De Houwer et al., 2020; Houwer, 2014; see also Kurdi & Dunham, 2020). Several studies demonstrated indeed that implicit and explicit measures are both shaped by factors that involve propositional reasoning (Kurdi & Dunham, 2020), posing limitations to the assumption of the dual-process theories according to which only explicit measures reflect the outcome of propositional processes.

Despite variations in the concepts put forth to account for dissociations between implicit and explicit measures, there is a shared agreement that the circumstances under which these instruments are used greatly influence their relationship and that the correlation between these measurements is stronger when they share similar contextual conditions (e.g.,

time constraints) and involve akin cognitive mechanisms (Gawronski et al., 2020; Gawronski & De Houwer, 2014).

Pervasiveness

Research using implicit measures showed that implicit attitudes and stereotypes are pervasive across demographic groups and topics (Guedj et al., 2021; Marini & Banaji, 2022; Nosek, Smyth, et al., 2007; Sabin et al., 2012). For example, a study testing a large sample of medical doctors (N=359,261) demonstrated that they hold strong implicit weight attitudes against overweight people similar to those of the general population (Sabin et al., 2012). Similarly, (Marini & Banaji, 2022) showed that STEM (science, technology, engineering, and math) faculty from the highest-ranked STEM universities in the U.S. exhibited an implicit belief associating women with sex than science as the general public. This belief was present both among female and male faculty. Notably, a large-scale study involving 2.5 million people also demonstrated that social psychological constructs measured using implicit measures are more pervasive than those assessed by means of explicit measures (Nosek, Smyth, et al., 2007). That is, for several topics (e.g., race, sexual orientation, age, and gender), people showed stronger implicit than explicit attitudes and stereotypes, although some exceptions were observed. For example, African Americans displayed stronger in-group attitudes at an explicit than an implicit level.

Predictivity

Over the past decades, several studies evaluated whether individual differences in implicit measures are predictive of variations in behavior. This research showed that implicit measures predict behaviors across a variety of topics and in many domains more effectively than explicit measures (Dovidio et al., 1997; Fazio et al., 1995; Greenwald, Poehlman, et al., 2009; Kurdi et al., 2019). For example, it has been shown that implicit measures predicted suicidal ideation and actual attempts (Nock et al., 2010), gender differences in science and math achievements across nations (Nosek et al., 2009), medical recommendations and treatments for patients (Green et al., 2007), hiring managers' decisions (Agerström & Rooth, 2011), and political voting behavior (Arcuri et al., 2008; Greenwald, Smith, et al., 2009).

Extensive research has explored the types of behaviors that implicit and explicit measures predict and the circumstances in which these instruments are effective in predicting behavior. Studies found that implicit measures are better at predicting spontaneous behavior (e.g., eye gaze in interracial interactions indicating racial prejudice), while explicit measures are more effective in predicting deliberate behavior (e.g., verbal responses indicating racial prejudice in interracial interactions) (Asendorpf et al., 2002; Dovidio et al., 2002; Fazio, 1990). In addition, it has been shown that the effectiveness of explicit and implicit measures in predicting behavior depends on the availability of processing resources. Explicit measures tend to be more predictive when processing resources are unconstrained, while implicit measures perform better when individuals have limited processing resources (Hofmann et al., 2007, 2008). For instance, Hofmann, Gschwendner, Castelli, and Schmitt (2008) found that interracial interactions were more strongly associated with implicit measures when participants were asked to perform a memory task. On the other hand, explicit measures were somewhat more predictive when participants were not engaged in the memory task and had full cognitive resources available.

The strength of the relationship between implicit measures and behavior also varies significantly depending on the topic and correspondence in measurement between the specific implicit and behavioral measures. For example, the predictive validity of implicit measures significantly exceeds that of explicit measures for socially sensitive topics, such as Black-White interracial behavior, while explicit measures tend to be more effective in predicting consumer and political preferences (Greenwald, Smith, et al., 2009; cf. Oswald et al., 2015). In addition, meta-analyses consistently find that the implicit-behavior relation is stronger when there is a higher correlation between implicit and explicit measures (Cameron et al., 2012; Greenwald, Poehlman, et al., 2009; Kurdi et al., 2019), supporting the idea that low implicit-explicit relations reflect a form of ambivalence that may hinder the prediction of behavior (Greenwald, Poehlman, et al., 2009; Lai & Wilson, 2021).

Recently, social scientists demonstrated that aggregated implicit data by area or state predicted the use of lethal force by police officers (Hehman et al., 2018) and even past slave distributions in the United States (Payne et al., 2019), highlighting the role of the social context in the relationship between implicit measures and behavior.

Malleability

A question of high relevance in implicit social cognition research is how we can produce changes in attitudes, stereotypes, and self-concepts measured using implicit measures.

Before 2000, social scientists thought implicit measures captured psychological constructs difficult to suppress (Macrae et al., 1994). Implicit attitudes, stereotypes, and self-concepts were believed to be habitual (Wilson et al., 2000), so engrained in the human mind that there was little that could be done to avoid their automatic activation and influence on behavior (Bargh, 1999). However, this belief was soon revised, showing that implicit social cognition could be shifted in response to brief interventions (Blair, 2002; Lai et al., 2014; Sritharan & Gawronski, 2010). For example, it has been demonstrated that implicit pro-White preferences (i.e., preferences for White people compared to Black people) could be even reversed after reading a counter-stereotypical scenario in which participants were asked to imagine themselves as victims of an assault and a White man and a Black man played the role of the aggressor and the rescuer, respectively (Lai et al., 2014; Marini et al., 2011; Marini, Rubichi, et al., 2012).

Over the past decades, several interventions have been designed to produce changes in implicit psychological constructs (Blair, 2002; Dasgupta & Greenwald, 2001; Gawronski & Bodenhausen, 2006; Lai et al., 2014; Sritharan & Gawronski, 2010). This research showed that implicit attitudes and stereotypes can be reduced by inducing changes in emotional states (DeSteno et al., 2004), exposing people to counter-stereotypical exemplars (Lai et al., 2014; Marini, Rubichi, et al., 2012), setting egalitarian goals (Legault et al., 2011; Mann & Kawakami, 2012), and providing educational programs (Kawakami et al., 2000; Rudman et al., 2001).

More recently, studies have shown that implicit attitudes and stereotypes can be shifted also using specific techniques - i.e., non-invasive brain stimulation, NBS (Marini et al., 2018), and virtual reality, VR (Tassinari et al., 2022) - aimed at changing their underlying cognitive mechanisms. Using transcranial direct-current stimulation (tDCS), a non-invasive brain-stimulation technique that produces changes in cortical activity, (Sellaro et al., 2015) showed that disruption of activity in the medial prefrontal cortex (mPFC) decreased implicit negative attitudes towards outgroup members. The mPFC is a brain area associated with the control and regulation processes that are involved in shaping, processing, and maintaining implicit attitudes and stereotypes, as well as in

their expression (Marini et al., 2018). Similar changes have been observed also by employing immersive Virtual Reality (VR). For example, it has been demonstrated that embodying White participants in a Black body reduced their implicit racial attitudes (Banakou et al., 2016; Marini & Casile, n.d.; Peck et al., 2013; Salmanowitz, 2018) and that this decrease was more likely to occur when the virtual environment had a positive valence (Banakou et al., 2020), the virtual social context was cooperative (Patané et al., 2020), or when participants could see their virtual body reflected in a mirror (Marini & Casile, n.d.).

Research has also demonstrated that intentional processes, such as people's focus attention, motivation to maintain a positive self-image and strategic efforts to reduce stereotypes, can moderate implicit stereotypes and attitudes (Blair, 2002). This suggests that once activated, implicit attitudes and stereotypes can be controlled. In line with these findings, Hahn and Gawronski (2019) found that directing people's attention to their implicit evaluations can increase awareness of their stereotypes and attitudes and subsequently reduce them.

Stability

Although research showed that implicit attitudes and stereotypes can be shifted in response to brief interventions, these changes were found to be not durable. For example, a study with a large sample size showed that even the most effective interventions to reduce implicit race attitudes (Lai et al., 2014; Marini, Rubichi, et al., 2012) produced only short-term changes that disappeared within few days (Lai et al., 2016). These findings suggest that implicit social cognition is relatively stable and that changing it may require extensive experience and time. In line with this view, Charlesworth and Banaji (2019) analyzed aggregate data over a 10-year period and showed that implicit attitudes can change in a stable and predictable manner. Specifically, they found that implicit sexuality attitudes (i.e., attitudes toward gay and straight people) revealed the largest change of any attitude analyzed in their study and were predicted to pass neutrality approximately in 2025. Further research is needed to understand the factors that may play a critical role in producing such changes over time.

Conclusion

The development of implicit measures has deeply influenced the field of social psychology and the study of intergroup relations. It showed, for the first time, that specific social psychological constructs considered in the past as inaccessible could be not only studied but also quantitatively measured, providing valuable insights for basic and applied research.

Studies using implicit measures allowed social and cognitive scientists to show that implicit attitudes and stereotypes are pervasive and relatively stable over time, even if they can be temporally shifted using specific techniques (i.e., NBS and VR) or behavioral interventions. Importantly, compared to explicit measures, they demonstrated to reveal complementary information and represent a better predictor of discriminatory behaviors in many domains. These findings produced relevant practical applications of implicit measures also in other disciplines and settings outside of the academic and research fields (Baron & Banaji, 2006; Craighero & Marini, 2021; Maison & Greenwald, 2001; Marini, Agosta, et al., 2012; Marini et al., 2016; Teachman et al., 2003). For example, in the healthcare setting, implicit measures have been employed to investigate medical disparities in treatment recommendations and communication behaviors (Hagiwara et al., 2020) and, in the domain of legal decision-making, to evaluate jury selection and sentencing decisions (Levinson & Smith, 2012).

Future research is however necessary to disentangle the mechanisms underlying implicit measures and translate these findings into the real world. Important questions and challenges are still open and unsolved. These include methodological aspects, interpretation of results, ambiguities in predicting behaviors better in some domains than others, as well as the lack of instruments with diagnostic abilities at a single-individual level (Gawronski et al., 2020; Greenwald & Lai, 2020; Lai & Wilson, 2021). What is certain, however, is that, independently from their setting of application or potential limitations, implicit measures helped to increase awareness of the existence of implicit prejudice and stereotyping and the development of training and interventions aimed at promoting integration and diversity in our society. Indeed, thanks to research using implicit measures, many corporations, organizations, schools, universities, and police departments now offer diversity training and programs in which people are educated about implicit attitudes and stereotypes and their potential consequences on behavior.

References

Arcuri, L., Castelli, L., Galdi, S., Zogmaister, C., & Amadori, A. (2008). Predicting the vote: implicit attitudes as predictors of the future behavior of decided and undecided voters. *Political Psychology*, *29*(3), 369–387. https://doi.org/10.1111/J.1467-9221.2008.00635.X

Asendorpf, J. B., Banse, R., & Mücke, D. (2002). Double dissociation between implicit and explicit personality self-concept: The case of shy behavior. In *Journal of Personality and Social Psychology* (Vol. 83, Issue 2). APA AMERICAN PSYCHOLOGICAL ASSOCIATION.

Banaji, M. R., & Hardin, C. D. (1996). Automatic stereotyping. *Psychological Science*, *7*(3), 136–141. https://doi.org/10.1111/J.1467-9280.1996.TB00346.X

Banakou, D., Beacco, A., Neyret, S., Blasco-Oliver, M., Seinfeld, S., & Slater, M. (2020). Virtual body ownership and its consequences for implicit racial bias are dependent on social context. *Royal Society Open Science*, *7*(12), 201848. https://doi.org/10.1098/RSOS.201848

Banakou, D., Hanumanthu, P. D., & Slater, M. (2016). Virtual embodiment of white people in a black virtual body leads to a sustained reduction in their implicit racial bias. *Frontiers in Human Neuroscience*, *10*, 601. https://doi.org/10.3389/fnhum.2016.00601

Bar-Anan, Y., & Nosek, B. A. (2014). A comparative investigation of seven indirect attitude measures. *Behavior Research Methods*, *46*(3), 668–688. https://doi.org/10.3758/s13428-013-0410-6

Bargh, J. A. (1999). The cognitive monster: the case against the controllability of automatic stereotype effects. In S. Chaiken & Y. Trope (Eds.), *Dual-process theories in social psychology* (pp. 361–382). The Guilford Press.

Bargh, J. A., & Williams, E. L. (2006). The automaticity of social life. *Current Directions in Psychological Science*, *15*(1), 1–4. https://doi.org/10.1111/J.0963-7214.2006.00395.X

Barnes-Holmes, D., Barnes-Holmes, Y., Power, P., Hayden, E., Milne, R., & Stewart, I. (2006). Do you really know what you believe? developing the implicit relational assessment procedure (IRAP) as a direct measure of implicit beliefs. *The Irish Psychologist*, *32*(7), 169–177.

Baron, A. S., & Banaji, M. R. (2006). The development of implicit attitudes. Evidence of race evaluations from ages 6 and 10 and adulthood. *Psychological Science*, *17*(1), 53–58.

Blair, I. V. (2002). The malleability of automatic stereotypes and prejudice. *Personality and Social Psychology Review*, *6*(3), 242–261.

Blair, I. V, & Banaji, M. R. (1996). Automatic and controlled processes in stereotype priming. *Journal of Personality and Social Psychology*, *70*(6), 1142–1163. https://doi.org/10.1037/0022-3514.70.6.1142

Blanton, H., & Jaccard, J. (2018). From principles to measurement: Theory-based tips on writing better questions. In H. Blanton, J. M. LaCroix, & G. D. Webster (Eds.), *Measurement in Social Psychology* (pp. 1–28). Routledge.

Bluemke, M., & Fiedler, K. (2009). Base rate effects on the IAT. *Consciousness and Cognition*, *18*(4), 1029–1038.

Cameron, C. D., Brown-Iannuzzi, J. L., & Payne, B. K. (2012). Sequential Priming Measures of Implicit Social Cognition. *Personality and Social Psychology Review*, *16*(4), 330–350. https://doi.org/10.1177/1088868312440047

Charlesworth, T. E. S., & Banaji, M. R. (2019). Patterns of implicit and explicit attitudes: I. Long-term change and stability from 2007 to 2016. *Psychological Science*, *30*(2), 174–192. https://doi.org/10.1177/0956797618813087

Chen, M., & Bargh, J. A. (1999). Consequences of automatic evaluation: immediate behavioral predispositions to approach or avoid the stimulus. *Personality and Social Psychology Bulletin*, *25*(2), 215–224. https://doi.org/10.1177/0146167299025002007

Craighero, L., & Marini, M. (2021). Implicit associations between adverbs of place and actions in the physical and digital space. *Brain Science*, *11*(11), 1523. https://doi.org/10.3390/brainsci11111523

Crowne, D. P., & Marlowe, D. (1960). A new scale of social desirability independent of psychopathology. *Journal of Consulting Psychology*, *24*(4), 349–354. https://doi.org/10.1037/H0047358

Cunningham, W. A., Nezlek, J. B., & Banaji, M. R. (2004). Implicit and explicit ethnocentrism: revisiting the ideologies of prejudice. *Personality and Social Psychology Bulletin*, *30*(10), 1332–1346. https://doi.org/10.1177/0146167204264654

Dasgupta, N., & Greenwald, A. G. (2001). On the malleability of automatic attitudes: Combating automatic prejudice with images of admired and disliked individuals. *Journal of Personality and Social Psychology*, *81*(5), 800.

De Houwer, J., Van Dessel, P., & Moran, T. (2020). *Attitudes beyond associations: On the role of propositional representations in stimulus evaluation* (pp. 127–183). https://doi.org/10.1016/bs.aesp.2019.09.004

DeSteno, D., Dasgupta, N., Bartlett, M. Y., & Cajdric, A. (2004). Prejudice from thin air: The effect of emotion on automatic intergroup attitudes.

Psychological Science : A Journal of the American Psychological Society / APS, *15*(5), 319–324. https://doi.org/10.1111/j.0956-7976.2004.00676.x

Dovidio, J. F., Kawakami, K., & Gaertner, S. L. (2002). Implicit and explicit prejudice and interracial interaction. *Journal of Personality and Social Psychology*, *82*(1), 62–68. https://doi.org/10.1037/0022-3514.82.1.62

Dovidio, J. F., Kawakami, K., Johnson, C., Johnson, B., & Howard, A. (1997). On the Nature of Prejudice: Automatic and Controlled Processes. *Journal of Experimental Social Psychology*, *33*(5), 510–540.

Eder, A. B., & Rothermund, K. (2008). When do motor behaviors (mis) match affective stimuli? An evaluative coding view of approach and avoidance reactions. *Journal of Experimental Psychology. General*, *137*(2), 262–281. https://doi.org/10.1037/0096-3445.137.2.262

Epifania, O. M., Anselmi, P., & Robusto, E. (2022). Implicit social cognition through the years: The Implicit Association Test at age 21. *Psychology of Consciousness: Theory, Research, and Practice*, *9*(3), 201–217. https://doi.org/10.1037/cns0000305

Fazio, R. H. (1990). *Multiple Processes by which Attitudes Guide Behavior: The Mode Model as an Integrative Framework* (pp. 75–109). https://doi.org/10.1016/S0065-2601(08)60318-4

Fazio, R. H. (2007). Attitudes as Object-Evaluation Associations of Varying Strength. *Social Cognition*, *25*(5), 603–637.

Fazio, R. H., Jackson, J. R., Dunton, B. C., & Williams, C. J. (1995). Variability in automatic activation as an unobtrusive measure of racial attitudes: a bona fide pipeline? *Journal of Personality and Social Psychology*, *69*(6), 1013–1027.

Fazio, R. H., Sanbonmatsu, D. M., Powell, M. C., & Kardes, F. R. (1986). On the automatic activation of attitudes. *Journal of Personality and Social Psychology*, *50*(2), 229–238. https://doi.org/10.1037/0022-3514.50.2.229

Gawronski, B., & Bodenhausen, G. V. (2006). Associative and propositional processes in evaluation: an integrative review of implicit and explicit attitude change. *Psychological Bulletin*, *132*(5), 692–731.

Gawronski, B., & De Houwer, J. (2014). Implicit measures in social and personality psychology. . In H. T. Reis & C. M. Judd (Eds.), *Handbook of research methods in social and personality psychology* (pp. 283–310). Cambridge University Press.

Gawronski, B., De Houwer, J., & Sherman, J. W. (2020). Twenty-five years of research using implicit measures. *Social Cognition*, *38*((Suppl)), S1–S25.

Greenwald, A. G., Brendl, M., Cai, H., Charlesworth, T., Cvencek, D., Dovidio. J. F., Friese, M., Hahn, A., Hehman, E., Hofmann, W., Hughes,

S., Hussey, I., Jordan, C., Jost, J., Kirby, T., Lai, C. K., Lang, J., Lindgren, K. P., Maison, D., … Wiers, R. W. (2019). The Implicit Association Test at age 20: What is known and what is not known about implicit bias. . *University of Washington. Retrieved from Https://Psyarxiv.Com/Bf97c.*

Green, A. R., Carney, D. R., Pallin, D. J., Ngo, L. H., Raymond, K. L., Iezzoni, L. I., & Banaji, M. R. (2007). Implicit bias among physicians and its prediction of thrombolysis decisions for black and white patients. *Journal of General Internal Medicine*, *22*(9), 1231–1238.

Greenwald, A. G., & Banaji, M. R. (1995). Implicit social cognition: attitudes, self-esteem, and stereotypes. *Psychological Review*, *102*(1), 4–27. https://doi.org/https://doi.org/10.1037/0033-295X.102.1.4

Greenwald, A. G., & Banaji, M. R. (2017). The implicit revolution: Reconceiving the relation between conscious and unconscious. *American Psychologist*, *72*(9), 861–871. https://doi.org/10.1037/amp0000238

Greenwald, A. G., & Lai, C. K. (2020). Implicit social cognition. *Annual Review of Psychology*, *71*, 419–445. https://doi.org/10.1146/ANNUREV-PSYCH-010419-050837

Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. (1998). Measuring individual differences in implicit cognition: the implicit association test. *Journal of Personality and Social Psychology*, *74*(6), 1464–1480. https://doi.org/https://doi.org/10.1037/0022-3514.74.6.1464

Greenwald, A. G., & Nosek, B. A. (2008). Attitudinal dissociation: What does it mean? In R. E. Petty, R. H. Fazio, & P. Briñol (Eds.), *Attitudes: Insights from the new implicit measures* (pp. 65–82). Psychology Press.

Greenwald, A. G., Poehlman, T. A., Uhlmann, E. L., & Banaji, M. R. (2009). Understanding and using the Implicit Association Test: III. Meta-analysis of predictive validity. *Journal of Personality and Social Psychology*, *97*(1), 17–41.

Greenwald, A. G., Smith, C. T., Sriram, N., Bar-Anan, Y., & Nosek, B. A. (2009). Implicit race attitudes predicted vote in the 2008 U.S. presidential election. *Analyses of Social Issues and Public Policy*, *9*(1), 241–253. https://doi.org/10.1111/J.1530-2415.2009.01195.X

Guedj, R., Marini, M., Kossowsky, J., Berde, C., Mateo, C., & Fleegler, E. W. (2021). Explicit and implicit bias based on race, ethnicity, and weight among pediatric emergency physicians. *Academic Emergency Medicine*, *28*(9), 1073–1076. https://doi.org/10.1111/ACEM.14301

Hagiwara, N., Dovidio, J. F., Stone, J., & Penner, L. A. (2020). Applied racial/ethnic Healthcare disparities research using implicit measures. *Social Cognition*, *38*(Suppl), S68–S97. https://doi.org/10.1521/SOCO.2020.38.

SUPP.S68

Hahn, A., & Gawronski, B. (2019). Facing one's implicit biases: From awareness to acknowledgment. *Journal of Personality and Social Psychology*, *116*(5), 769–794. https://doi.org/10.1037/pspi0000155

Hehman, E., Flake, J. K., & Calanchini, J. (2018). Disproportionate use of lethal force in policing is associated with regional racial biases of residents. *Social Psychological and Personality Science*, *9*(4), 393–401. https://doi.org/10.1177/1948550617711229

Hofmann, W., Gschwendner, T., Castelli, L., & Schmitt, M. (2008). Implicit and Explicit Attitudes and Interracial Interaction: The Moderating Role of Situationally Available Control Resources. *Group Processes & Intergroup Relations*, *11*(1), 69–87. https://doi.org/10.1177/1368430207084847

Hofmann, W., Rauch, W., & Gawronski, B. (2007). And deplete us not into temptation: Automatic attitudes, dietary restraint, and self-regulatory resources as determinants of eating behavior. *Journal of Experimental Social Psychology*, *43*(3), 497–504. https://doi.org/10.1016/j.jesp.2006.05.004

Houwer, J. De. (2014). A Propositional Model of Implicit Evaluation. *Social and Personality Psychology Compass*, *8*(7), 342–353. https://doi.org/10.1111/spc3.12111

Jacoby, L. L. (1991). A process dissociation framework: Separating automatic from intentional uses of memory. *Journal of Memory and Language*, *30*(5), 513–541. https://doi.org/10.1016/0749-596X(91)90025-F

Karpinski, A., & Steinman, R. B. (2006). The single category implicit association test as a measure of implicit social cognition. *Journal of Personality and Social Psychology*, *91*(1), 16–32. https://doi.org/10.1037/0022-3514.91.1.16

Kawakami, K., Dovidio, J. F., Moll, J., Hermsen, S., & Russin, A. (2000). Just say no (to stereotyping): effects of training in the negation of stereotypic associations on stereotype activation. *Journal of Personality and Social Psychology*, *78*(5), 871–888. https://doi.org/10.1037//0022-3514.78.5.871

Krieglmeyer, R., & Deutsch, R. (2010). Comparing measures of approach–avoidance behaviour: The manikin task vs. two versions of the joystick task. *Cognition & Emotion*, *24*(5), 810–828. https://doi.org/10.1080/02699930903047298

Kurdi, B., & Dunham, Y. (2020). Propositional accounts of implicit evaluation: Taking stock and looking ahead. *Social Cognition, Supplement*, 43–68.

Kurdi, B., Seitchik, A. E., Axt, J. R., Carroll, T. J., Karapetyan, A., Kaushik, N., Tomezko, D., Greenwald, A. G., & Banaji, M. R. (2019). Relationship between the Implicit Association Test and intergroup behavior: A meta-analysis. *American Psychologist*, *74*(5), 569–586.

Lai, C. K., Marini, M., Lehr, S. A., Cerruti, C., Shin, J.-E. L., Joy-Gaba, J. A., Ho, A. K., Teachman, B. A., Wojcik, S. P., Koleva, S. P., Frazier, R. S., Heiphetz, L., Chen, E. E., Turner, R. N., Haidt, J., Kesebir, S., Hawkins, C. B., Schaefer, H. S., Rubichi, S., … Nosek, B. A. (2014). Reducing Implicit Racial Preferences: I. A Comparative Investigation of 17 Interventions. *Journal of Experimental Psychology: General*, *143*(4), 1765–1785.

Lai, C. K., Skinner, A. L., Cooley, E., Murrar, S., Brauer, M., Devos, T., Calanchini, J., Xiao, Y. J., Pedram, C., Marshburn, C. K., Simon, S., Blanchar, J. C., Joy-Gaba, J. A., Conway, J., Redford, L., Klein, R. A., Roussos, G., Schellhaas, F. M. H., Burns, M., … Nosek, B. A. (2016). Reducing implicit racial preferences: II. Intervention effectiveness across time. *Journal of Experimental Psychology. General*, *145*(8), 1001–1016.

Lai, C. K., & Wilson, M. E. (2021). Measuring implicit intergroup biases. *Social and Personality Psychology Compass*, *15*(1), e12573. https://doi.org/10.1111/SPC3.12573

Legault, L., Gutsell, J. N., & Inzlicht, M. (2011). Ironic effects of antiprejudice messages: how motivational interventions can reduce (but also increase) prejudice. *Psychological Science*, *22*(12), 1472–1477. https://doi.org/10.1177/0956797611427918

Levinson, J. D., & Smith, R. J. (2012). Implicit racial bias across the law. In *Cambridge University Press*. Cambridge University Press. https://doi.org/10.1017/CBO9780511820595

Maass, A., Ceccarelli, R., & Rudin, S. (1996). Linguistic intergroup bias: evidence for in-group-protective motivation. *Journal of Personality and Social Psychology*, *71*(3), 512–526. https://doi.org/10.1037/0022-3514.71.3.512

Maass, A., Milesi, A., Zabbini, S., & Stahlberg, D. (1995). Linguistic intergroup bias: differential expectancies or in-group protection? *Journal of Personality and Social Psychology*, *68*(1), 116–126. https://doi.org/10.1037//0022-3514.68.1.116

Maass, A., Salvi, D., Arcuri, L., & Semin, G. R. (1989). Language use in intergroup contexts: the linguistic intergroup bias. *Journal of Personality and Social Psychology*, *57*(6), 981–993. https://doi.org/10.1037//0022-3514.57.6.981

Macrae, C. N., Bodenhausen, G. V., Milne, A. B., & Jetten, J. (1994). Out of mind but back in sight: stereotypes on the rebound. *Journal of Personality and Social Psychology*, *67*(5), 808–817. https://doi.org/10.1037/0022-3514.67.5.808

Maison, D., & Greenwald, A. G. (2001). The Implicit Association Test as a measure of implicit consumer attitudes. *Polish Psychological Bulletin*, *32*(1), 1–9.

Mann, N. H., & Kawakami, K. (2012). The long, steep path to equality: progressing on egalitarian goals. *Journal of Experimental Psychology: General*, *141*(1), 187–197. https://doi.org/10.1037/A0025602

Marini, M. (2017). Underweight vs. overweight/obese: which weight category do we prefer? Dissociation of weight-related preferences at the explicit and implicit level. *Obesity Science & Practice*, *3*(4), 390–398.

Marini, M., Agosta, S., Mazzoni, G., Dalla Barba, G., & Sartori, G. (2012). True and false DRM memories: differences detected with an implicit task. *Frontiers in Psychology*, *3*, 310.

Marini, M., Agosta, S., & Sartori, G. (2016). Electrophysiological Correlates of the Autobiographical Implicit Association Test (aIAT): Response Conflict and Conflict Resolution. *Frontiers in Human Neuroscience*, *10*, 391.

Marini, M., & Banaji, M. R. (2022). An implicit gender sex-science association in the general population and STEM faculty. *Journal of General Psychology*, *149*(3), 299–326. https://doi.org/10.1080/00221309.2020.1853502

Marini, M., Banaji, M. R., & Pascual-Leone, A. (2018). Studying implicit social cognition with noninvasive brain stimulation. *Trends in Cognitive Sciences*.

Marini, M., & Casile, A. (n.d.). I can see my virtual body in a mirror: the role of visual perspective in changing implicit racial attitudes using VR. *Frontiers in Psychology*.

Marini, M., Rubichi, S., & Sartori, G. (2011). Implicit racial stereotypes may temporarily shift after reading a story. In L. Carlson, C. Hoelscher, & T. Shipley (Eds.), *33rd Annual Conferences of Cognitive Science Society* (pp. 1224–1229). Cognitive Science Society.

Marini, M., Rubichi, S., & Sartori, G. (2012). The role of self-involvement in shifting IAT effects. *Experimental Psychology*, *59*(6), Art. 0.

Marini, M., Sriram, N., Schnabel, K., Maliszewski, N., Devos, T., Ekehammar, B., Wiers, R., HuaJian, C., Somogyi, M., Shiomura, K., Schnall, S., Neto, F., Bar-Anan, Y., Vianello, M., Ayala, A., Dorantes, G., Park, J., Kesebir, S., Pereira, A., … Nosek, B. A. (2013). Overweight

people have low levels of implicit weight bias, but overweight nations have high levels of implicit weight bias. *PloS One*, *8*(12), e83543. https://doi.org/10.1371/journal.pone.0083543

Marini, M., Waterman, P. D., Breedlove, E., Chen, J. T., Testa, C., Reisner, S. L., Pardee, D. J., Mayer, K. H., & Krieger, N. (2021). The target/perpetrator brief-implicit association test (B-IAT): an implicit instrument for efficiently measuring discrimination based on race/ethnicity, sex, gender identity, sexual orientation, weight, and age. *BMC Public Health*, *21*(1). https://doi.org/10.1186/s12889-021-10171-7

Marini, M., Waterman, P. D., Breedlove, E. R., Chen, J. T., Testa, C., Pardee, D. J., LeBlanc, M., Reisner, S. L., Oendari, A., & Krieger, N. (2022). Using implicit measures of discrimination: White, Black, and Hispanic participants respond differently to group-specific racial/ethnic categories vs. the general category "People of Color" in the USA. *Journal of Racial and Ethnic Health Disparities*. https://doi.org/10.1007/S40615-022-01353-Z

Nock, M. K., Park, J. M., Finn, C. T., Deliberto, T. L., Dour, H. J., & Banaji, M. R. (2010). Measuring the Suicidal Mind. *Psychological Science*, *21*(4), 511–517. https://doi.org/10.1177/0956797610364762

Nosek, B. A. (2007). Implicit-explicit relations. *Current Directions in Psychological Science*, *16*(2), 65–69. https://doi.org/https://doi.org/10.1111/j.1467-8721.2007.00477.x

Nosek, B. A., Banaji, M., & Greenwald, A. G. (2002). Harvesting implicit group attitudes and beliefs from a demonstration web site. *Group Dynamics: Theory, Research, and Practice*, *6*(1), 101–115.

Nosek, B. A., & Banaji, M. R. (2001). The GO/NO-GO Association Task. *Social Cognition*, *19*(6), 625–664. https://doi.org/10.1521/SOCO.19.6.625.20886

Nosek, B. A., Greenwald, A. G., & Banaji, M. R. (2007). The Implicit Association Test at 7 age 7: A methodological and conceptual review. In J. A. Bargh (Ed.), *Automatic processes in social thinking and behavior* (pp. 265–292).

Nosek, B. A., Hawkins, C. B., & Frazier, R. S. (2011). Implicit social cognition: from measures to mechanisms. *Trends in Cognitive Sciences*, *15*(4), 152–159.

Nosek, B. A., Smyth, F. L., Hansen, J. J., Devos, T., Linder, N. M., Ranganath, K. A., Smith, C. T., Olson, K. R., Chugh, D., Greenwald, A. G., & Banaji, M. R. (2007). Pervasiveness and correlates of implicit attitudes and stereotypes. *European Review of Social Psychology*, *18*, 36–88.

Nosek, B. A., Smyth, F. L., Sriram, N., Lindner, N. M., Devos, T., Ayala,

A., Bar-Anan, Y., Bergh, R., Cai, H., Gonsalkorale, K., Kesebir, S., Maliszewski, N., Neto, F., Olli, E., Park, J., Schnabel, K., Shiomura, K., Tulbure, B. T., Wiers, R. W., … Greenwald, A. G. (2009). National differences in gender-science stereotypes predict national sex differences in science and math achievement. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(26), 10593–10597. https://doi.org/https://doi.org/10.1073/pnas.0809921106

Nosek, B. A., Sriram, N., Smith, C. T., & Bar-Anan, Y. (2014). *The Multi-Category Implicit Association Test.*

Olivola, C. Y., Funk, F., & Todorov, A. (2014). Social attributions from faces bias human choices. *Trends in Cognitive Sciences*, *18*(11), 566–570. https://doi.org/10.1016/J.TICS.2014.09.007

Oswald, F. L., Mitchell, G., Blanton, H., Jaccard, J., & Tetlock, P. E. (2015). Using the IAT to predict ethnic and racial discrimination: Small effect sizes of unknown societal significance. *Journal of Personality and Social Psychology*, *108*(4), 562–571. https://doi.org/10.1037/pspa0000023

Paladino, M. P., & Castelli, L. (2008). On the immediate consequences of intergroup categorization: activation of approach and avoidance motor behavior toward ingroup and outgroup members. *Personality & Social Psychology Bulletin*, *34*(6), 755–768. https://doi.org/10.1177/0146167208315155

Patané, I., Lelgouarch, A., Banakou, D., Verdelet, G., Desoche, C., Koun, E., Salemme, R., Slater, M., & Farnè, A. (2020). Exploring the effect of cooperation in reducing implicit racial bias and its relationship with dispositional empathy and political attitudes. *Frontiers in Psychology*, *11*. https://doi.org/10.3389/FPSYG.2020.510787

Payne, B. K., Cheng, C. M., Govorun, O., & Stewart, B. D. (2005). An inkblot for attitudes: affect misattribution as implicit measurement. *Journal of Personality and Social Psychology*, *89*(3), 277–293. https://doi.org/10.1037/0022-3514.89.3.277

Payne, B. K., Vuletich, H. A., & Brown-Iannuzzi, J. L. (2019). Historical roots of implicit bias in slavery. *Proceedings of the National Academy of Sciences of the United States of America*, *116*(24), 11693–11698. https://doi.org/10.1073/pnas.1818816116

Payne, K., & Lundberg, K. (2014). The Affect Misattribution Procedure: ten years of evidence on reliability, validity, and mechanisms. *Social and Personality Psychology Compass*, *8*(12), 672–686. https://doi.org/10.1111/SPC3.12148

Peck, T. C., Seinfeld, S., Aglioti, S. M., & Slater, M. (2013). Putting yourself in the skin of a black avatar reduces implicit racial bias.

Consciousness and Cognition, *22*(3), 779–787. https://doi.org/10.1016/j.concog.2013.04.016

Reingold, E. M., & Merikle, P. M. (1988). Using direct and indirect measures to study perception without awareness. *Perception & Psychophysics*, *44*(6), 563–575. https://doi.org/10.3758/BF03207490

Richardson-Klavehn, A., & Bjork, R. A. (1988). Measures of Memory. *Annual Review of Psychology*, *39*(1), 475–543. https://doi.org/10.1146/annurev.ps.39.020188.002355

Rudman, L. A., Ashmore, R. D., & Gary, M. L. (2001). "Unlearning" automatic biases: the malleability of implicit prejudice and stereotypes. *Journal of Personality and Social Psychology*, *81*(5), 856–868. https://doi.org/10.1037//0022-3514.81.5.856

Rydell, R. J., & McConnell, A. R. (2006). Understanding implicit and explicit attitude change: A systems of reasoning analysis. *Journal of Personality and Social Psychology*, *91*(6), 995–1008. https://doi.org/10.1037/0022-3514.91.6.995

Sabin, J. A., Marini, M., & Nosek, B. A. (2012). Implicit and explicit anti-fat bias among a large sample of medical doctors by gender, BMI and race/ethnicity. *PLoS ONE*, *7*(11), e48448.

Salmanowitz, N. (2018). The impact of virtual reality on implicit racial bias and mock legal decisions. *Journal of Law and the Biosciences*, *5*(1), 174–203. https://doi.org/10.1093/JLB/LSY005

Sellaro, R., Derks, B., Nitsche, M. A., Hommel, B., van den Wildenberg, W. P. M., van Dam, K., & Colzato, L. S. (2015). Reducing prejudice through brain stimulation. *Brain Stimulation*, *8*(5), 891–897. https://doi.org/10.1016/j.brs.2015.04.003

Sidanius, J., Pratto, F., Martin, M., & Stallworth, L. M. (1991). Consensual Racism and Career Track: Some Implications of Social Dominance Theory. *Political Psychology*, *12*(4), 691. https://doi.org/10.2307/3791552

Sriram, N., & Greenwald, A. G. (2009). The brief implicit association test. *Experimental Psychology*, *56*(4), 283–294. https://doi.org/10.1027/1618-3169.56.4.283

Sritharan, R., & Gawronski, B. (2010). Changing implicit and explicit prejudice: Insights from the associative-propositional evaluation model . *Social Psychology*, *41*(3), 113–123.

Strack, F., & Deutsch, R. (2004). Reflective and Impulsive Determinants of Social Behavior. *Personality and Social Psychology Review*, *8*(3), 220–247. https://doi.org/10.1207/s15327957pspr0803_1

Tassinari, M., Aulbach, M. B., & Jasinskaja-Lahti, I. (2022). The use of virtual reality in studying prejudice and its reduction: A systematic review.

PloS One, *17*(7). https://doi.org/10.1371/JOURNAL.PONE.0270748

Teachman, B. A., Gapinski, K. D., Brownell, K. D., Rawlins, M., & Jeyaram, S. (2003). Demonstrations of implicit anti-fat bias: the impact of providing causal information and evoking empathy. *Health Psychology*, *22*(1), 68–78.

Wentura, D., & Degner, J. (2010). A practical guide to sequential priming and related tasks. In B. Gawronski & B. K. Payne (Eds.), *Handbook of implicit social cognition: measurement, theory, and applications* (pp. 95–116). The Guilford Press.

Wheatley, T., & Wegner, D. M. (2001). Automaticity of action, psychology of. *International Encyclopedia of the Social & Behavioral Sciences*, 991–993. https://doi.org/10.1016/B0-08-043076-7/01747-2

Wilson, T. D., Lindsey, S., & Schooler, T. Y. (2000). A model of dual attitudes. *Psychological Review*, *107*(1), 101–126. https://doi.org/10.1037/0033-295X.107.1.101

Wittenbrink, B., Judd, C. M., & Park, B. (1997). Evidence for racial prejudice at the implicit level and its relationship with questionnaire measures. *Journal of Personality and Social Psychology*, *72*(2), 262–274. https://doi.org/10.1037//0022-3514.72.2.262

18. When social variables shape social attention: The case of ethnic group membership

Mario Dalmaso[1], Giovanni Galfano[1], Luigi Castelli[1]
[1]University of Padova, Italy

People tend to shift their visual attention towards the same spatial location gazed at by other individuals, a phenomenon known as gaze cueing of attention. This phenomenon was initially considered as automatic, in that it appeared to be relatively insensitive to manipulations that are used to test different flavours of automaticity such as expectancies. However, in recent years, an increasing number of studies has outlined the conditionally automatic nature of this effect. Indeed, gaze cueing can be also shaped by several social variables characterising 1) the observer, 2) the face providing the gaze cue, and 3) their relationship. The present chapter focuses on the third category and summarises the main findings concerning the possible modulatory role of ethnic group membership on gaze cueing.

Gaze cueing of attention

During social interactions, we tend to shift our visual attention towards the same location as that indicated by averted-gaze faces (Frischen et al., 2007). This form of social attention is known as gaze cueing of attention and allows us to establish meaningful and fluid interactions with both others and the environment around us (Emery, 2000). From an experimental perspective, gaze cueing of attention has been largely investigated by using the so-called gaze-cueing task (e.g., Driver et al., 1999;

Friesen & Kingstone, 1998) in which, typically, a central face is presented with the gaze averted either leftward or rightward (see also Figure 1). Then, a peripheral target appears, and participants are asked to provide a response (e.g., a key press) as soon as they notice the target. Importantly, gaze direction is task-irrelevant in that it is not predictive of the spatial location of the upcoming target. If the target appears in the same spatial location gazed at by the face, a better performance (i.e., shorter latencies and greater accuracy) is observed compared to when the target appears elsewhere (that is, the gaze-cueing effect). Another version of this task, known as the oculomotor interference paradigm, is based on eye movements rather than manual responses (e.g., Dalmaso et al., 2020b; Kuhn & Kingstone, 2009; Ricciardelli et al., 2002). During the task, a central face is presented with the gaze averted either leftward or rightward and, on each trial, participants receive the instruction to perform a saccadic eye movement towards a left or a right placeholder. Also in this case, a better performance typically emerges when the direction of the instructed saccade and the direction of the gaze stimulus match (e.g., right-right) as compared to when they differ (e.g., right-left), even if gaze direction is irrelevant for the task.

A variety of studies on the gaze-cueing effect mainly used schematic faces (which are particularly suited for controlling low-level visual properties; see also Figure 1) and reported that the gaze-cueing effect is a strong and reflexive phenomenon (Driver et al., 1999; Galfano et al., 2012). This means that every time we observe a face with an averted gaze, our attentional system would tend to produce an attentional shift towards the location suggested by the eye gaze. Although this pattern seems to hold in the case of schematic faces, we also must keep in mind that, during real-life social interactions, we typically meet individuals characterised by a variety of social variables such as gender, age, familiarity, and ethnicity. In complex social environments, populated by individuals differing with respect to several group memberships and features, selection processes may intervene in order to prioritise the stimuli that are expected to provide the most relevant information in a given context (e.g., ingroup members; leaders). Thus, it seems reasonable to expect that, according to this functional perspective, the gaze-cueing effect could be modulated as a function of the perceived informativeness and social meaning associated with the faces due to their group memberships. This assumption has been confirmed by a flourishing literature and a summary of the major findings in this field can be found in a recent review (Dal-

maso et al., 2020). More details about the impact of social variables on gaze cueing are provided in the next paragraph.
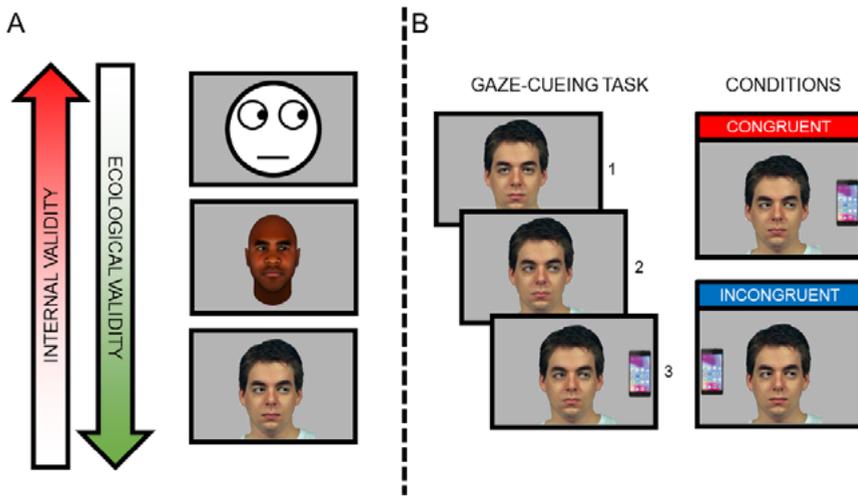
Figure 1. Panel A illustrates examples of stimuli used to study gaze cueing of attention. For example, they can be schematic faces, avatars, or photographs of real individuals. These stimuli are characterised by different levels of ecological validity (lower for schematic faces, intermediate for avatars, higher for real faces) and internal validity (higher for schematic faces, intermediate for avatars, and lower for real faces). Panel B illustrates a possible trial used in a standard gaze cueing task. The central face is presented with a direct gaze and then with the gaze averted rightward. Then, a target object (here, the picture of a smartphone) appears in a spatially congruent location, namely in the same location gazed at by the central face. The gaze-cueing effect is computed by comparing performance (e.g., mean response latency) in congruent and incongruent trials.

The social side of gaze cueing

One way to systematise the impact of social variables on the gaze-cueing effect is to cluster them into three main categories: 1) the characteristics of the observer, 2) the characteristics of the cueing face (i.e., the face providing the gaze stimulus) and 3) the relationship between the characteristics of the observer and those of the cueing face (Dalmaso et al., 2020a). As for the first category, it has been observed that, for instance, female individuals tend to present a greater gaze-cueing effect as compared to male individuals (e.g., Bayliss et al., 2005), in line with the idea that females would be more sensitive to social signals than males

(e.g., Geary, 1998). As for the second category, it has been observed, for instance, that faces expressing negative emotions (e.g., anger) tend to elicit a greater gaze-cueing effect as compared to faces expressing positive emotions (e.g., happiness), in line with the notion that our attentional system would be particularly sensitive to negative stimuli, as they could signal the presence of dangerous stimuli in the environment (e.g., Pecchinenda et al., 2008). As for the third category, a series of studies reported that the gaze-cueing effect can be modulated according to the interaction between the ethnic group membership of the observer and that of the cueing face. These studies will be illustrated in the next paragraph.

The role of ethnic membership in shaping gaze cueing of attention

The first study exploring the possible role of ethnic group membership on gaze cueing presented White and Black faces to White and Black individuals (Pavan et al., 2011). All participants were tested in Italy as students at the University of Padova. The results showed that while a reliable gaze-cueing effect emerged in Black participants in response to both White and Black faces, White participants showed a reliable gaze-cueing effect only in response to White faces. This latter pattern of results has been then replicated also in a subsequent study employing the oculomotor interference paradigm (Dalmaso et al., 2015), in which White individuals (Italian students at the University of Padova) again showed to be more influenced by eye-gaze stimuli provided by White faces than by Black faces.

Importantly, a conceptual replication of these studies has also been provided by a study (Weisbuch et al., 2017) conducted in a different socio-cultural context (i.e., the United States). As in Pavan et al. (2011), Black individuals showed gaze cueing of attention for both White and Black faces, whereas White individuals showed gaze cueing of attention only for White faces. Interestingly, an additional experiment also reported that the gaze-cueing effect for Black faces was however detectable among White participants for whom a feeling of low power was induced (i.e., after imagining that others have power over them). The results reported by Weisbuch et al. (2017) are of great interest, as they do suggest that perceived social hierarchies would be involved in the modulation of gaze cueing based on ethnic group membership.

More recently, a study by Zhang and colleagues (Zhang, Dalmaso, Castelli, Fiorese, et al., 2021a) investigated the gaze-cueing effect by

comparing White individuals living in a Western country (i.e., Italy) and Asian individuals living in an Eastern country (i.e., China). In this case, the main results showed that while for White individuals the gaze-cueing effect was reliable and similar in magnitude for both White and Asian faces – in line with a previous study conducted in the United Kingdom among British participants (Strachan et al., 2017) – Asian individuals showed to be much more influenced by the gaze provided by White faces rather than by Asian faces. Zhang et al. (2021a) also collected self-report measures about the perceived social status associated with White and Asian individuals. The results showed that while White respondents reported perceiving the social status of White and Asian people as rather similar, Asian respondents reported perceiving the social status of White people as significantly higher than the social status of Asian people. This pattern, which reminds of the findings reported by Weisbuch et al. (2017), can be interpreted as further evidence supporting the idea that social hierarchies may be the underlying factor driving the modulation of gaze cueing based on ethnic group membership. The results reported by Zhang et al. (2021a) have been then replicated also in an oculomotor interference study (Zhang, Dalmaso, Castelli, Fu, et al., 2021b) showing that, also in this case, eye movements of White individuals tested in Italy were similarly influenced by gaze stimuli provided by White and Asian faces, while eye movements of Asian individuals tested in China were more influenced by gaze stimuli belonging to White faces than Asian faces.

Finally, a recent study conducted in China (Zhang et al., 2023), replicated the pattern described above, namely a larger gaze-cueing effect for facial stimuli depicting White rather than Asian individuals. However, this pattern of results only emerged when the two types of faces were presented intermixed within the task (i.e., in each trial, a White or an Asian face could be randomly presented). Critically, when the two types of faces were presented into two distinct blocks of trials (i.e., in one block only White faces appeared, in another block only Asian faces appeared), the gaze-cueing effect was not modulated by face type. The rationale is that the presentation in an intermixed fashion of faces belonging to different ethnic groups would favour the activation of categorization processes based on a continuous comparison of different types of stimuli. In contrast, such comparison is, by definition, lacking when stimuli belonging to different ethnic groups are presented in a blocked fashion (see Macrae & Cloutier, 2009). Overall, these results suggest that gaze-cueing of attention can be shaped by social

variables associated with the face stimuli, but also that such modulation is context-dependent and it emerges only when social comparison processes are triggered. This points to the relevance of selection processes that are crucial to navigate in complex social environments and that may operate by leading to more efficient responses to stimuli that are appraised as the likely bearers of critical information in a given context (e.g., ingroup members; high-status individuals).

Conclusion

In this chapter, we have briefly summarised the main findings showing that gaze cueing of attention, namely the tendency to shift our visual attention in response to others' eye-gaze direction, can be shaped by ethnic group membership. The available evidence suggests that social hierarchies play a key role in the emergence of such modulations since individuals tend to be less influenced by faces belonging to social groups associated with a lower social status. This conclusion is also supported by other works on gaze cueing – not based on ethnic group membership – in which the social status and dominance of the presented faces were directly manipulated (Dalmaso et al., 2012, 2014; Jones et al., 2010; Ohlsen et al., 2013). Also in these cases, a greater gaze cueing can be observed in response to highly dominant faces than to less dominant faces, likely because dominant individuals are more likely associated with a greater amount of resources and, more generally, following the cues they provide could lead to greater benefits than following the cues from less dominant or submissive individuals (Jones et al., 2010).

To conclude, the gaze-cueing task can be interpreted as a simple and powerful tool for studying human behaviour in social contexts. The impact of ethnic group membership on gaze cueing is of particular relevance and could represent a fruitful avenue towards which to push future research, given the increasing degree of multiculturality which characterises modern societies and because it remains a relatively unexplored topic in social cognition.

References

Bayliss, A. P., di Pellegrino, G., & Tipper, S. P. (2005). Sex differences in eye gaze and symbolic cueing of attention. The Quarterly Journal

of Experimental Psychology Section A, 58(4), 631–650. https://doi.org/10.1080/02724980443000124

Dalmaso, M., Castelli, L., & Galfano, G. (2020a). Social modulators of gaze-mediated orienting of attention: A review. Psychonomic Bulletin & Review, 27(5), 833–855. https://doi.org/10.3758/s13423-020-01730-x

Dalmaso, M., Castelli, L., & Galfano, G. (2020b). Early saccade planning cannot override oculomotor interference elicited by gaze and arrow distractors. Psychonomic Bulletin & Review, 27(5), 990–997. https://doi.org/10.3758/s13423-020-01768-x

Dalmaso, M., Galfano, G., & Castelli, L. (2015). The Impact of Same- and Other-Race Gaze Distractors on the Control of Saccadic Eye Movements. Perception, 44(8–9), 1020–1028. https://doi.org/10.1177/0301006615594936

Dalmaso, M., Galfano, G., Coricelli, C., & Castelli, L. (2014). Temporal Dynamics Underlying the Modulation of Social Status on Social Attention. PLoS ONE, 9(3), e93139. https://doi.org/10.1371/journal.pone.0093139

Dalmaso, M., Pavan, G., Castelli, L., & Galfano, G. (2012). Social status gates social attention in humans. Biology Letters, 8(3), 450–452. https://doi.org/10.1098/rsbl.2011.0881

Driver, J., Davis, G., Ricciardelli, P., Kidd, P., Maxwell, E., & Baron-Cohen, S. (1999). Gaze perception triggers reflexive visuospatial orienting. Visual Cognition, 6(5), 509–540. https://doi.org/10.1080/135062899394920

Emery, N. J. (2000). The eyes have it: the neuroethology, function and evolution of social gaze. Neuroscience & Biobehavioral Reviews, 24(6), 581–604. https://doi.org/10.1016/S0149-7634(00)00025-7

Friesen, C. K., & Kingstone, A. (1998). The eyes have it! Reflexive orienting is triggered by nonpredictive gaze. Psychonomic Bulletin & Review, 5(3), 490–495. https://doi.org/10.3758/BF03208827

Frischen, A., Bayliss, A. P., & Tipper, S. P. (2007). Gaze cueing of attention: visual attention, social cognition, and individual differences. Psychological Bulletin, 133(4), 694–724. https://doi.org/10.1037/0033-2909.133.4.694

Galfano, G., Dalmaso, M., Marzoli, D., Pavan, G., Coricelli, C., & Castelli, L. (2012). Eye gaze cannot be ignored (but neither can arrows). The Quarterly Journal of Experimental Psychology, 65(10), 1895–1910. https://doi.org/10.1080/17470218.2012.663765

Geary, D. C. (1998). Male, female: The evolution of human sex differences. American Psychological Association. https://doi.org/10.1037/10370-000

Jones, B. C., DeBruine, L. M., Main, J. C., Little, A. C., Welling, L. L. M., Feinberg, D. R., & Tiddeman, B. P. (2010). Facial cues of dominance modulate the short-term gaze-cuing effect in human observers. Proceedings. Biological Sciences / The Royal Society, 277(1681), 617–624. https://doi.org/10.1098/rspb.2009.1575

Kuhn, G., & Kingstone, A. (2009). Look away! Eyes and arrows engage oculomotor responses automatically. Attention, Perception, and Psychophysics, 71(2), 314–327. https://doi.org/10.3758/APP.71.2.314

Macrae, C. N., & Cloutier, J. (2009). A matter of design: Priming context and person perception. Journal of Experimental Social Psychology, 45(4), 1012–1015. https://doi.org/10.1016/j.jesp.2009.04.021

Ohlsen, G., van Zoest, W., & van Vugt, M. (2013). Gender and Facial Dominance in Gaze Cuing: Emotional Context Matters in the Eyes That We Follow. PLoS ONE, 8(4), e59471. https://doi.org/10.1371/journal.pone.0059471

Pavan, G., Dalmaso, M., Galfano, G., & Castelli, L. (2011). Racial group membership is associated to gaze-mediated orienting in Italy. PLoS ONE, 6(10), e25608. https://doi.org/10.1371/journal.pone.0025608

Pecchinenda, A., Pes, M., Ferlazzo, F., & Zoccolotti, P. (2008). The combined effect of gaze direction and facial expression on cueing spatial attention. Emotion (Washington, D.C.), 8(5), 628–634. https://doi.org/10.1037/a0013437

Ricciardelli, P., Bricolo, E., Aglioti, S., & Chelazzi, L. (2002). My eyes want to look where your eyes are looking: Exploring the tendency to imitate another individual's gaze. Neuroreport, 13(17). https://doi.org/10.1097/01.wnr.0000044227

Strachan, J. W. A. A., Kirkham, A. J., Manssuer, L. R., Over, H., & Tipper, S. P. (2017). Incidental learning of trust from eye-gaze: Effects of race and facial trustworthiness. Visual Cognition, 25(7–8), 802–814. https://doi.org/10.1080/13506285.2017.1338321

Weisbuch, M., Pauker, K., Adams, R. B., Lamer, S. A., & Ambady, N. (2017). Race, power, and reflexive gaze following. Social Cognition, 35(6), 619–638. https://doi.org/10.1521/soco.2017.35.6.619

Zhang, X., Dalmaso, M., Castelli, L., Fiorese, A., Lan, Y., Sun, B., Fu, S., & Galfano, G. (2021a). Social attention across borders: A cross-cultural investigation of gaze cueing elicited by same- and other-ethnicity faces. British Journal of Psychology, 112(3), 741–762. https://doi.org/10.1111/bjop.12476

Zhang, X., Dalmaso, M., Castelli, L., Fu, S., & Galfano, G. (2021b). Cross-cultural asymmetries in oculomotor interference elicited by gaze

distractors belonging to Asian and White faces. Scientific Reports, 11(1), 20410. https://doi.org/10.1038/s41598-021-99954-x

Zhang, X., Dalmaso, M., Galfano, G., & Castelli, L. (2023). Tuning social modulations of gaze cueing via contextual factors. https://doi.org/10.3758/s13423-022-02211-z

19. Environment learning, gender differences, and the role of spatial beliefs

Francesca Pazzaglia[1], Chiara Meneghetti[1], Laura Miola[1], Veronica Muffato[1]
University of Padova, Italy

Space, navigation, and environment learning

The space surrounding people is an essential part of their daily behaviours and cognition because people live in and are constantly surrounded by space (Ishikawa & Zhou, 2020). Navigating within an environment relies on different processes, such as locomotion and wayfinding (Montello, 2001). Locomotion is based on sensory information to guide individuals' movements and requires coordination accessible to the sensory and motor systems. Wayfinding consists in the goal-directed behaviour and relies on individuals' mental representations, planning, decision making, and spatial reasoning.

During navigation, people simultaneously learn spatial information, such as landmarks, distances between landmarks, directions, and the environment's path networks, forming the so-called spatial knowledge (see Montello, 2001; Montello & Raubal, 2013). Learning spatial information from the environment leads to forming a *cognitive map*, a concept Tolman (1948) first introduced and was subsequently conceived of as a flexible (e.g., not associated to a specific orientation) mental representation of the environment (Wolbers & Hegarty, 2010).

The quality and the features of the environment representation can be assessed indirectly using different spatial recall tasks. Some tasks measure egocentric (observer based) knowledge (also called route knowledge)

in which the person's point of view is a reference to completing the task. For example, repeating a previously navigated route or identifying the order in which the person encountered landmarks represents egocentric tasks. Other tasks assess allocentric (viewpoint independent) knowledge (also called survey knowledge) in which landmarks and locations need to be related to each other to accomplish the task. For example, finding a shortcut in a previously navigated environment or drawing a map of the environment are allocentric tasks that require identifying connections between landmarks and locations.

People differ widely in performing recall tasks that assess environment information, which suggests a large variability in navigation ability and acquiring spatial knowledge (Montello, 1998; Wolbers & Hegarty, 2010). Even commonly, in everyday spatial situations, each individual knows someone who is good at navigation and others who are less so.

Therefore, a deep investigation of the relationship between individual factors and environment learning and navigation is important. The next sections describe some individual factors likely to be crucial in spatial learning and navigation, such as gender, visuospatial abilities, and beliefs related to navigation ability.

Gender

Spatial cognition is a cognitive domain that showed gender differences in favour of men (Halpern, 2012). For example, a recent meta-analysis analyzing various studies showed that, overall, men outperformed women in navigation ability and environment learning (Nazareth et al., 2019). Similarly, another systematic review found men's better performance in both spatial navigation (large-scale abilities) and visuospatial abilities (small-scale abilities; Yuan et al., 2019).

However, despite the overall general difference in favour of men, notably, results between studies were inconsistent and seem to depend on the type of tasks used to test spatial knowledge (for a review, see Coluccia & Louse, 2004). For example, after learning an environment from navigation, there was evidence that men outperformed women in survey tasks (e.g., Boone et al., 2018) while no systematic differences emerged in route tasks (e.g., Castelli et al., 2008).

In addition to navigation and environment learning, some of the most studied gender differences concern mental rotation, such as the ability to imagine how an object will appear when rotated. Several studies and

reviews of gender differences showed that men outperformed women across all age ranges (Geiser et al., 2008; Moè, 2018; Voyer et al., 1995), while other studies suggested attenuated male and female differences in aging due to cognitive age decline (Borella et al., 2014; Jansen & Heil, 2009).

There may be many factors behind differences in spatial abilities and researchers have proposed some explanations. Gender differences can refer to different strategies men and women use, with women preferring to encode landmark information from an egocentric first-person perspective and men focusing more on the environment's configural global information (Coluccia & Louse, 2004; Saucier et al., 2002). Other explanations proposed different levels of spatial abilities or an influence of hormone levels; finally, different experiences occurring throughout life may enhance one's spatial abilities (Coluccia & Louse, 2004; Lawton, 2010, Boone et al., 2018).

However, to date, most of the studies on gender differences in spatial abilities and their explanations focused especially on cognitive aspects and accounted for a part of the results. We believe that holding a comprehensive view of individual differences considering personal beliefs is necessary to better understand gender differences and individual differences.

Spatial abilities

One of the main sources of individual differences in environment learning consists in the ability to generate, retain, and process abstract visual image (Lohman, 1988), that is, spatial abilities. These abilities are high-order cognitive ones that Linn and Petersen (1985) classified in three factors: spatial perception (the ability to determine spatial relations with respect to one's own position), spatial visualization (the ability to perform manipulations of complex spatial information), and mental rotation (the ability to rotate figures mentally). Spatial abilities also encompass maintaining and processing visuospatial information (visuospatial working memory; Logie, 1995), and it has been well demonstrated that high-order processing contribute to sustaining environment learning and acquiring spatial information (Hegarty et al., 2006; Meneghetti et al., 2014; Meneghetti et al., 2021).

Beliefs about navigation ability: Self-efficacy, growth mindset, and gender stereotypes

Other sources of variability that likely explain gender differences in spatial domain are individual beliefs, such as self-efficacy, growth mindset, and stereotypes.

Spatial Self-efficacy

Bandura (1994) defined self-efficacy as people's beliefs about their capabilities to produce levels of performance. Self-efficacy is not a general evaluation about personal ability, but rather is tied to a specific domain and comprises judging personal capabilities to perform given task demands. Self-efficacy can be considered a hierarchical set of beliefs (Herzog & Dixon, 1994), from more global (e.g., "I have a good memory") to more task-specific beliefs (e.g., "I can remember these objects"; Bandura, 1989; Beaudoin & Desrichard, 2011). Central to the self-efficacy theory is that self-efficacy and competence are directly associated with actual behaviours and performance. Self-efficacy promotes a person's performance through the personal strategies, effort, persistence, and affect (Bandura, 1977).

Furthermore, an external way to promote self-efficacy includes reinforcement given in terms of normative feedbacks, that is, giving information on one's compared to others' performance (for example above-average performance). Normative feedbacks can have the power to boost performance in various tasks, from arithmetic to name recall and motor skills learning (e.g., Strickland-Hughes et al., 2017; Wulf et al., 2010), being an effective intervention to promote cognitive and motor performance (Peifer et al., 2020).

We applied self-efficacy to the spatial domain, referring to "spatial self-efficacy," defined as personal beliefs of one's ability to accomplish environmental and navigation tasks (Mitolo et al., 2015; Pazzaglia et al., 2017; Pazzaglia et al., 2018). Spatial self-efficacy is typically associated with the ability to find a shortcut and to other navigation tasks (Pazzaglia et al., 2017; Pazzaglia et al., 2018).

Spatial growth mindset

Another belief related to spatial abilities is *growth mindset*. In general, a growth (or incremental) mindset refers to a set of beliefs regarding

the possibility to improve and enhance an ability. A growth mindset is opposed to a fixed (or entity) one, in which ability is considered not malleable, and consequently, improvable (Dweck, 2006).

The growth mindset has been widely investigated with respect to academic achievement and intelligence. Adopting a growth mindset of intelligence (i.e., considering intelligence as improvable through effort and commitment) positively influences student achievement (Mueller & Dweck, 1998), resulting in the tendency to persist through failure, set more adaptive goals, and engage in challenging tasks (Dweck, 2006).

He and Hegarty (2020) compared the growth mindset towards intelligence and navigation abilities, finding that people were more likely to consider their intelligence improvable than their spatial navigation. In addition, they showed that the growth mindset was positively associated with navigation ability. In our studies described below, we found that the growth mindset was relevant to promoting functional navigation behaviours (Miola et al., under review).

However, these studies used self-report measures for assessing navigation behaviours and navigation ability, thus, further research on objective performance is needed to better understand the role of the growth mindset in navigation ability.

Stereotypes

Stereotypes consist of a set of beliefs likely to influence individuals' behaviours and performances. Gender stereotypes, which entail beliefs about behaviours or characteristics of each sex (Del Boca & Ashmore, 1980), have been investigated when considering spatial ability, in particular mental rotation, with inconsistent results (e.g., Guizzo et al., 2019; Moè 2012; for a meta-analysis, see also Doyle & Voyer, 2016).

Fewer studies have investigated gender stereotypes with respect to spatial navigation. Crawford (1989) found that both genders considered men better than women in recalling places and directions. More recently, one study found that men (but not women) improved their navigation after receiving instructions, activating a comparison with the opposite gender (Rosenthal et al., 2012). Finally, an effect of a stereotype threat emerged for both men and women, but only in a strongly engaging task (Allison et al., 2017).

Taken together, the existing literature on beliefs related to spatial abilities is promising in suggesting their relationships with spatial tasks.

The following section is dedicated to further exploring the mechanisms of such relations.

The relationship between beliefs about spatial abilities, environment learning, and navigation behaviours

In the present section, we describe three recent studies conducted in our lab that examined the mutual relationships between beliefs about spatial abilities (i.e., self-efficacy, growth mindset, gender stereotype), environment learning, and navigation behaviours (i.e., GPS use and pleasure in exploring). All these studies explored the role of spatial self-efficacy in environment learning and navigation, but in two studies, we considered global and task-specific self-efficacy in explaining environment learning and gender differences in environment learning; whereas in another study, we specifically examined whether and how the growth mindset and gender stereotypes in the context of navigation interacted with self-efficacy in predicting navigation behaviours.

In the first study (Miola et al., 2021), participants were asked to actively navigate through a route within a virtual environment and successively indicate the starting point (pointing task) and locate landmarks on a sketch map of the environment (map-location task). Before performing each task, they were required to evaluate their confidence with the task (task-specific self-efficacy). Global spatial self-efficacy and spatial abilities were also assessed at the beginning of the experiment. Using a structural equation model, we could examine the mutual relationship between gender, spatial abilities, (global and task specific) spatial self-efficacy, and the two spatial tasks (map-location task and pointing task). Interestingly, we found that gender differences in performing the map-completion task were mediated by spatial abilities and (global and task specific) spatial self-efficacy, suggesting the importance of considering such variables in explaining gender differences in spatial domain.

On these bases, in another study (Miola et al., 2021), we combined spatial abilities and self-efficacy with an experimental manipulation, implementing normative feedback with two participant groups having, respectively, positive versus neutral fictitious feedbacks on their performance of visuospatial tasks. Successively, they navigated through a virtual route and then performed three spatial tasks (route repetition, pointing, and map location). We also collected their confidence with each task (task-specific self-efficacy) and their global self-efficacy. Unexpectedly,

no differences emerged between the positive and neutral feedback, but results from mediation showed that receiving feedback was indirectly related to the performance in all three spatial-recall tasks through task-specific self-efficacy. In other words, receiving positive feedback related to higher levels of self-efficacy that, in turn, predicted performance in spatial tasks.

Finally, a third study (Miola et al., under revision) involved 609 participants who responded to questionnaires on spatial self-efficacy, the growth mindset, and gender stereotypes related to spatial and navigation ability. All these measures were put in relation with two opposite navigation behaviours—the tendency of exploring the environment and that of using GPS. The results showed that the growth mindset positively related to self-efficacy in both men and women, suggesting that the higher people trust in the possibility to improve their ability, the more self-efficacy is perceived during spatial tasks. Then, we found an opposite relation between gender stereotypes and self-efficacy in men and women. The belief that men were better than women (which was found in the whole sample) negatively related to self-efficacy in women and positively in men. Finally, women referred lower spatial self-efficacy and exploration tendency and a higher use of GPS than men did.

Conclusions

To conclude, in the present overview, we outlined studies on individual differences regarding beliefs about spatial abilities and environment learning. Overall, these findings suggest that self-efficacy as well as other beliefs related to personal spatial ability (the growth mindset and stereotypes) can relate to individuals' spatial performance and behaviours. Moreover, there is promising evidence that what really matters is not gender *per se*, but rather women's and men's beliefs in their spatial and navigation skills. As a direct consequence, programs to improve spatial abilities and orientation in women and men should integrate cognitive trainings with motivational boosters.

We are particularly grateful to Anne Maass because an initial collaboration with her inspired this new perspective that has successively been expanded. Thank you, Anne!

References

Allison, C., Redhead, E. S., & Chan, W. (2017). Interaction of task difficulty and gender stereotype threat with a spatial orientation task in a virtual nested environment. *Learning and Motivation*, *57*, 22-35. https://doi.org/10.1016/j.lmot.2017.01.005

Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review*, *84,* 191-215. https://doi.org/10.1037/0033-295X.84.2.191

Bandura, A. (1989). Regulation of cognitive processes through perceived self-efficacy. *Developmental Psychology*, *25*, 729–735. doi:10.1037/0012-1649.25.5.729

Bandura, A. (1994). Self-efficacy. In V. S. Ramachandran (Eds.), *Encyclopedia of Human Behavior* (Vol. 4, pp. 71–81). Academic Press.

Beaudoin, M., & Desrichard, O. (2011). Are memory self-efficacy and memory performance related? A meta-analysis. *Psychological Bulletin*, *137*, 211-241. https://doi.org/10.1037/a0022106

Boone, A. P., Gong, X., & Hegarty, M. (2018). Sex differences in navigation strategy and efficiency. *Memory & cognition*, *46*, 909-922. https://doi.org/10.3758/s13421-018-0811-y

Borella, E., Meneghetti, C., Ronconi, L., & De Beni, R. (2014). Spatial abilities across the adult life span. *Developmental psychology*, *50*, 384-392. https://doi.org/10.1037/a0033818

Castelli, L., Corazzini, L. L., & Geminiani, G. C. (2008). Spatial navigation in large-scale virtual environments: Gender differences in survey tasks. *Computers in Human Behavior*, *24*, 1643-1667. https://doi.org/10.1016/j.chb.2007.06.005

Coluccia, E., & Louse, G. (2004). Gender differences in spatial orientation: A review. *Journal of Environmental Psychology*, *24*, 329-340. https://doi.org/10.1016/j.jenvp.2004.08.006

Crawford, M., Herrmann, D. J., Holdsworth, M. J., Randall, E. P., & Robbins, D. (1989). Gender and beliefs about memory. *British Journal of Psychology*, *80*(3), 391-401. https://doi.org/10.1111/j.2044-8295.1989.tb02329.x

Del Boca, F. K., & Ashmore, R. D. (1980). Sex stereotypes and implicit personality theory. II. A trait-inference approach to the assessment of sex stereotypes. *Sex Roles*, *6*(4), 519-535. https://doi.org/10.1007/BF00287883

Doyle, R. A., & Voyer, D. (2016). Stereotype manipulation effects on math and spatial test performance: A meta-analysis. *Learning and Individual*

Differences, *47*, 103-116. https://doi.org/10.1016/j.lindif.2015.12.018

Dweck, C. S. (2006). *Mindset: The new psychology of success*. Random House.

Geiser, C., Lehmann, W., & Eid, M. (2008). A note on sex differences in mental rotation in different age groups. I*ntelligence*, *36*, 556-563. https://doi.org/10.1016/j.intell.2007.12.003

Guizzo, F., Moè, A., Cadinu, M., & Bertolli, C. (2019). The role of implicit gender spatial stereotyping in mental rotation performance. *Acta psychologica*, *194*, 63-68. https://doi.org/10.1016/j.actpsy.2019.01.013

Halpern, D. F. (2012). Sex differences in cognitive abilities (4th ed.). Hove, UK: Psychology Press.

He, C., & Hegarty, M. (2020). How anxiety and growth mindset are linked to navigation ability: Impacts of exploration and GPS use. *Journal of Environmental Psychology*, *71*, 101475. https://doi.org/10.1016/j.jenvp.2020.101475

Hegarty, M., Montello, D. R., Richardson, A. E., Ishikawa, T., & Lovelace, K. (2006). Spatial abilities at different scales: Individual differences in aptitude-test performance and spatial-layout learning. *Intelligence*, *34*, 151-176. https://doi.org/10.1016/j.intell.2005.09.005

Hertzog, C., & Dixon, R. A. (1994). Metacognitive development in adulthood and old age. In J. Metcalfe & A. P. Shimamura (Eds.), *Metacognition: Knowing about knowing* (pp. 227–251). MIT Press.

Ishikawa, T., & Zhou, Y. (2020). Improving cognitive mapping by training for people with a poor sense of direction. *Cognitive Research: Principles and Implications*, *5*, 1-19. https://doi.org/10.1186/s41235-020-00238-1

Jansen, P., & Heil, M. (2009). Gender differences in mental rotation across adulthood. *Experimental aging research*, *36*, 94-104. https://doi.org/10.1080/03610730903422762

Lawton, C. A. (2010). Gender, spatial abilities, and wayfinding. In *Handbook of gender research in psychology* (pp. 317-341). Springer. https://doi.org/10.1007/978-1-4419-1465-1_16

Linn, M. C., & Petersen, A. C. (1985). Emergence and characterization of sex differences in spatial ability: A meta-analysis. *Child Development*, *56*, 1479-1498. https://doi.org/10.2307/1130467.

Logie, R. H. (1995). Visuo-spatial working memory. Hove, UK: L. Erlbaum.

Lohman, D. F. (1988). Spatial abilities as traits, processes, and knowledge. In R. J. Sternberg (Ed.), *Advances in the psychology of human intelligence* (pp. 181–248). Lawrence Erlbaum Associates, Inc.

Meneghetti, C., Borella, E., Pastore, M., & De Beni, R. (2014). The role of spatial abilities and self-assessments in cardinal point orientation

across the lifespan. *Learning and Individual Differences*, *35*, 113-121. https://doi.org/10.1016/j.lindif.2014.07.006

Meneghetti, C., Miola, L., Toffalini, E., Pastore, M., & Pazzaglia, F. (2021). Learning from navigation, and tasks assessing its accuracy: The role of visuospatial abilities and wayfinding inclinations. *Journal of Environmental Psychology*, *75*, 101614. https://doi.org/10.1016/j.jenvp.2021.101614

Miola, L., Meneghetti, C., Toffalini, E., & Pazzaglia, F. (2021). Environmental learning in a virtual environment: Do gender, spatial self-efficacy, and visuospatial abilities matter? *Journal of Environmental Psychology*, *78*, 101704. https://doi.org/10.1016/j.jenvp.2021.101704

Miola, L., Muffato, V., Meneghetti, C., & Pazzaglia, F. (2021). Spatial Learning in a Virtual Environment: The Role of Self-Efficacy Feedback and Individual Visuospatial Factors. Brain Sciences, 11(9), 1185. https://doi.org/10.3390/brainsci11091185

Mitolo, M., Gardini, S., Caffarra, P., Ronconi, L., Venneri, A., & Pazzaglia, F. (2015). Relationship between spatial ability, visuospatial working memory and self-assessed spatial orientation ability: a study in older adults. *Cognitive Processing*, *16*, 165-176. https://doi.org/10.1007/s10339-015-0647-3

Moè, A. (2012). Gender difference does not mean genetic difference: Externalizing improves performance in mental rotation. *Learning and Individual differences*, *22*, 20-24. https://doi.org/10.1016/j.lindif.2011.11.001

Moè, A. (2018). Mental rotation and mathematics: Gender-stereotyped beliefs and relationships in primary school children. *Learning and Individual Differences*, *61*, 172-180. https://doi.org/10.1016/j.lindif.2017.12.002

Montello D. R. (2001). Spatial cognition. In *International encyclopedia of the social & behavioral sciences* (pp. 14771–14775). Oxford: Pergamon Press.

Montello, D. R. (1998). A new framework for understanding the acquisition of spatial knowledge in largescale environments. In M. J. Egenhofer & R. G. Golledge (Eds.), *Spatial and temporal reasoning in geographic information system* (pp. 143-154). Oxford University Press.

Montello, D., & Raubal, M. (2013). Functions and applications of spatial cognition. In D. Waller & L. Nadel (Eds.), *The APA handbook of spatial cognition*, (pp. 249-264). American Psychology Association. https://doi.org/10.1037/13936-014

Mueller, C. M., & Dweck, C. S. (1998). Praise for intelligence can undermine

children's motivation and performance. *Journal of personality and social psychology*, *75*, 33. https://doi.org/10.1037/0022-3514.75.1.33

Nazareth, A., Huang, X., Voyer, D., & Newcombe, N. (2019). A meta-analysis of sex differences in human navigation skills. *Psychonomic Bulletin & Review*, *26*, 1503-1528. https://doi.org/10.3758/s13423-019-01633-6

Pazzaglia, F., Meneghetti, C., & Ronconi, L. (2018). Tracing a route and finding a shortcut: The working memory, motivational, and personality factors involved. *Frontiers in Human Neuroscience*, *12*, 225. https://doi.org/10.3389/fnhum.2018.00225

Pazzaglia, F., Meneghetti, C., Labate, E., & Ronconi, L. (2017). Are wayfinding self-efficacy and pleasure in exploring related to shortcut finding? A study in a virtual environment. In Spatial Cognition X (pp. 55-68). Springer, Cham. https://doi.org/10.1007/978-3-319-68189-4_4

Peifer, C., Schönfeld, P., Wolters, G., Aust, F., & Margraf, J. (2020). Well done! Effects of positive feedback on perceived self-efficacy, flow and performance in a mental arithmetic task. *Frontiers in Psychology.*, 1008. https://doi.org/10.3389/fpsyg.2020.010

Rosenthal, H. E., Norman, L., Smith, S. P., & McGregor, A. (2012). Gender-based navigation stereotype improves men's search for a hidden goal. *Sex roles*, *67*, 682-695. https://doi.org/10.1007/s11199-012-0205-8

Saucier, D. M., Green, S. M., Leason, J., MacFadden, A., Bell, S., & Elias, L. J. (2002). Are sex differences in navigation caused by sexually dimorphic strategies or by differences in the ability to use the strategies? *Behavioral Neuroscience*, *116*, 403–410. https://doi.org/10.1037/0735-7044.116.3.403

Strickland-Hughes, C. M., West, R. L., Smith, K. A., & Ebner, N. C. (2017). False feedback and beliefs influence name recall in younger and older adults. *Memory*, *25*(8), 1072–1088. https://doi.org/10.1080/09658211.2016.1260746

Tolman, E. C. (1948). Cognitive maps in rats and men. *Psychological Review*, *55*, 189–208.

Wolbers, T., & Hegarty, M. (2010). What determines our navigational abilities? *Trends in Cognitive Sciences*, *14*, 138-146. https://doi.org/10.1016/j.tics.2010.01.001

Wulf, G., Chiviacowsky, S., & Lewthwaite, R. (2010). Normative feedback effects on learning a timing task. *Research Quarterly for Exercise and Sport*, *81*, 425–431. https://doi.org/10.1080/02701367.2010.10599703

Yuan, L., Kong, F., Luo, Y., Zeng, S., Lan, J., & You, X. (2019). Gender differences in large-scale and small-scale spatial ability: A systematic

review based on behavioral and neuroimaging research. *Frontiers in Behavioral Neuroscience*, *13*, 128. https://doi.org/10.3389/fnbeh.2019.00128

20. How stereotypes fuel academic disparities: the stereotype threat model

Silvia Galdi
University of Campania, Italy

The 21st century has brought with it unparallelled levels of diversity in the classroom. It is now common to see students of different ethnicity, gender, or religious affiliation at all levels of education. This change held out the hope of elimination of inequalities in career advancement and academic opportunities. This expectation, however, has not been met yet.

Academic disparities have been a matter of great concern among social scientists, policymakers, and the general public for a very long time. Standard explanations typically invoke either nature (i.e., biological and genetic explanations) or nurture (i.e., explanations based on environment, culture, and socialization). For example, Herrnstein and Murray (1994) claimed that members of some social groups underachieve in educational settings because they are genetically endowed with inferior intelligence. Similar suggestions have been made to explain women's under-representation in math and science (Bembow & Stanley, 1980), with Harvard University President Lawrence Summers' allusion to women's inferior "intrinsic aptitude" being only the most high-profile of references to this possibility (Summers, 2005).

On the nurture side of the debate is the view that some combination of environmental factors hinders members of some social groups from developing the appropriate skills, values, and motivation needed for success. For example, being raised in a low-income family often means having less access to educational resources in addition to limited access to health care and nutrition, both of which contribute to lower academic performance (Croizet & Millet, 2012). Cultural and socialization pressures

may also contribute to the gap (e.g., Tenenbaum & Leaper, 2003). Mothers, for example, are more likely to encourage their sons than daughters to work hard in math and science, despite evidence indicating that their daughters perform equally well, if not better than their sons in these domains (Scafidi & Bui, 2010).

Whether due to biology or accrued effects of upbringing, these explanations share the presumption that people belonging to some social groups have less ability. But the theory that is the focus of the present chapter assumes that even if we could match students on genetic predispositions, educational and social background, and personal values, something in the situation itself may hold some people back from reaching their full potential. This something is stereotypes.

Social psychologists Claude Steele and Joshua Aronson posited that the mere salience of stereotypes might create extra pressure for those who are negatively stereotyped in a specific area of competence, an experience that is burdensome enough to systematically impair, in the short term, their ability to perform in a task on that domain up to their potential. Steele and Aronson's hunch was soon confirmed. In their first experiment, Steele and Aronson (1995) had African American and White college students take a challenging test on verbal reasoning. In the diagnostic test condition, they presented the test as a measure of intellectual ability. In the non-diagnostic test condition, they removed the relevance of the negative stereotype about African Americans (alleging their inferior intellectual abilities) and said that they just wanted to use the test to examine the psychology of verbal problem-solving. This was the only difference between the two conditions of the experiment: the test was the same, students were equally talented, and received the same amount of time. However, this little difference in the way the test was presented made a big difference for the African American students: they underperformed on the verbal reasoning test when it was presented as a diagnostic indicator of their intellectual ability. On the contrary, when the same test was presented as non-diagnostic of their ability, they performed equivalently to their White peers. In other words, Steele and Aronson demonstrated that performance can be spoiled by conditions that make ability stereotypes relevant and improved by conditions that nullify them. These results were revolutionary. Steele and Aronson shifted the focus from the nature vs. nurture trap as a source of achievement gaps, to an effect that was both external and immediate: the momentary impact of stereotypes, which is what the two authors called *stereotype threat*.

In this chapter, I will provide a brief overview of the stereotype threat model. The following section introduces the boundary conditions of stereotype threat. Next, I will examine some moderating variables that help to explain why some people are more vulnerable to the stereotype threat than others. Subsequently, I will address the question of what processes drive stereotype threat. The final section will be dedicated to the potential long-term effects of stereotype threat.

Stereotype threat: necessary conditions

According to the stereotype threat model (Steele, 1997; Steele & Aronson, 1995), when tested on a domain in which one's social group is stereotyped as a poor performer (e.g., African Americans testing in intellectual ability; women testing in math) and situational cues signal that the stereotype is relevant as a possible interpretation for one's behavior, one becomes concerned that anything they say or do could be interpreted through the lens of low stereotypic expectations. This apprehensiveness about confirming the ingroup ability stereotype ironically leads to impaired performance. Of course, stereotype threat is one causal factor, but not the sole cause that makes stereotyped individuals underperform. Two meta-analyses (Nguyen & Ryan 2008; Nadler & Clark 2011) corroborate the assumption that stereotype threat explains some - but not all - of the mean group differences in cognitive tests. Nonetheless, over the past three decades, more than three hundred experimental studies have illustrated the deleterious and extensive effects that stereotype threat can inflict on many different populations and ages ranging from children (e.g., Galdi et al., 2014) to seniors (e.g., Maass & Cadinu, 2003).

Originally developed to explain racial differences (Whites versus Blacks) in academic performance, stereotype threat has been subsequently used to shed light on disparities between men and women in mathematical problem-solving (Cadinu et al., 2005), people of high and low socio-economic status on linguistic tasks (Croizet & Claire, 1998), and old and young adults on short-term memory (Mazerolle et al., 2012), among others. Other studies have demonstrated that the possibility of confirming an ability stereotype about one's social group affects performance not only in academic domains, but also in non-academic domains as diverse as negotiations (Kray et al. 2002), financial decision-making (Carr & Steele 2010), golf putting (Stone et al. 1999), driving (Skorich et al., 2013; Yeung & von Hippel,

2008), childcare (Bosson et al., 2004), doctor-patient interactions (Burgess et al., 2010), and workplace success (Gupta et al., 2014).

Research has also uncovered information about the conditions necessary for stereotype threat to occur. Now we know that for stereotypes to affect performance, one must i) *self-categorize* as a member of a *group negatively stereotyped in a given domain*; ii) be *aware* of the stereotype; iii) be in a situation where the stereotype is *salient* and *relevant* as a possible interpretation for performance; iv) while engaging in a *difficult* task on the stereotyped domain.

The most basic criterion is simply identifying themselves as members of a social group negatively stereotyped in a given domain, such that the self is defined in terms of group membership (i.e., *self-categorization*). Since each person has multiple social identities (e.g., gender, age, ethnicity, socioeconomic status, etc.; Tajfel and Turner, 1986), this implies that *every* individual is potentially vulnerable to stereotype threat, because every social group exhibits at least one ability-impugning stereotype in some given situation. For example, White men, a social group that has a relatively positive social status, underperform when they believe that their math performance will be compared to that of Asian men (Aronson et al., 1999); White men also perform worse than Black men when a motor task is related to "natural athletic ability" (Stone, 2002; Stone et al., 1999). Therefore, people from all social groups, including those who do not belong to traditionally stigmatized groups, can experience stereotype threat.

The second primary criterion, *stereotype awareness,* requires that the stereotype exists and the target is aware of it. That is, people within a society must have a shared schema or belief about members of particular social groups, and members of those groups must know that the ingroup stereotype is culturally broadly held. Thus, although believing that a given ability stereotype about one's group is true (referred to as stereotype endorsement) can exacerbate the threat experience (Huguet & Régner, 2009; Wheeler & Petty, 2001), in order to be affected by stereotype threat people need *only know* that others may endorse and apply the ability stereotype about one of their social identities to them.

The third core requirement is that experiences of stereotype threat are rooted in an environment's situational cues that make a stereotype currently salient and relevant to one's actions. Stereotype threat is indeed fundamentally driven by something in the moment - any situational cue - indicating that an individual is at risk of being judged in the light of an

ability stereotype related to one of their social identities. Situational cues that foster *stereotype salience* include verbally reminding participants of the ingroup ability stereotype prior to a test of their performance in the relevant domain (e.g., Spencer et al., 1999), informing that the task measures a stereotyped skill (e.g., by explicitly labeling tasks "math test" or "intelligence test"), or simply soliciting identity-relevant information (e.g., asking students for their demographic information, including gender and ethnicity) before taking the test (Danhaer & Crandall, 2008).

Also the physical arrangements and mere presence of certain social groups within a setting, such as being outnumbered by members of non-stereotyped social groups (Ben-Zeev et al. 2004, Inzlicht & Good 2004) or being taught by an instructor who is a member of a non-stereotyped outgroup (Marx & Roman 2002), may represent subtle but powerful situational cues that foster the stereotyped group identity to become more salient. Other research has demonstrated just how insidious and commonplace situational cues can be. For example, being exposed to media images or interacting with others may be sufficient to trigger stereotype threat, as women showed reduced math performance after they interacted with people displaying sexist behavior (Logel et al., 2009), as well as after exposure to gender-stereotypic advertisements (Davies et al. 2002, 2005). Therefore, stereotype threat effects can be obtained by simply reminding targets of culturally held stereotypes or by emphasizing a stereotyped group identity. Moreover, given that people tend to be highly sensitive to cues indicating that one of their social identities might be devalued (Purdie-Vaughns et al. 2008, Steele et al. 2002, Wout et al. 2009), cues do not have to be necessarily blatant.

It is also important to note that for stereotype threat to occur the stereotype must be not only salient but also relevant to evaluating performance. A striking example of this criterion comes from a study by Cadinu, Maass, Lombardo, and Frigerio (2006), who manipulated the test label and, hence, its *stereotype relevance*. A group of men and women were given the same test that was presented as measuring either logical intelligence (the stereotype relating to women entails that they have low ability in the STEM field) or social intelligence (the stereotype relating to men entails that they have poor social intelligence). Results showed that women performed worse when the test allegedly measured logical rather than social intelligence, whereas the opposite was found for men. Because the decrease in performance was only observed in the domain in which a given group is negatively stereotyped, this study provides strong

evidence that stereotype threat occurs specifically in stereotype-relevant domains.

The final primary criterion is the evaluation of performance in the stereotyped domain, thus creating the opportunity to confirm or disconfirm the stereotype. Typically, the valuation takes place in test or exam contexts where the individual is concerned about confirming the stereotype in the eyes of other people (Shapiro & Neuberg, 2007). However, stereotype threat effects have been found even when the performance is totally private (e.g., Inzlicht & Ben-Zeev, 2003; Shapiro, 2011; Wout et al., 2008), demonstrating that it is the *possibility of evaluation* in the stereotyped domain *per se*, irrespective of the audience, that is necessary for stereotype threat to occur. Research has also shown that the task at hand must be *difficult* in order to elicit debilitating stereotype threat effects. Concerns about confirming an ability stereotype with their performance, indeed, make threatened individuals try harder on tests in order to invalidate the stereotype (Aronson, 2002). As we will see in the section about processes underlying stereotype threat, this increase in effort and anxiety can be an advantage in situations where effort or a rush of adrenaline is desirable, such as on easy or well-learned tasks where more effort pays off (O'Brien & Crandall, 2003). Conversely, on difficult tasks a sort of relaxed concentration is critical, and anything that compounds performance pressure represents a handicap (Aronson, 2002).

For whom stereotype threat influences outcomes

Although stereotype threat is a robust and widespread phenomenon, people are not equally susceptible to its debilitating effects. To date, a number of individual difference variables (i.e., moderating factors) that render individuals more or less vulnerable to stereotype threat have been identified. Many variables can be interpreted as heightening (vs. lowering) one of the basic criteria for stereotype threat effects, such as self-stereotyping. For example, individuals who are highly identified with the stereotyped ingroup (i.e., for whom ingroup membership is more accessible; Cadinu & Galdi, 2012) show increased vulnerability to stereotype threat compared to those less identified (Schmader, 2002; Wout et al., 2008). For example, only women who were highly identified with their gender performed worse than men on a math test that was described as evaluating the abilities of women in general (Schmader, 2002). Other moderators enhance (vs. reduce) the salience or relevance of the stereotype. Research

has demonstrated that those who are invested in the stereotyped domain (e.g., a woman highly motivated to pursue math) are more vulnerable because they care most about doing well and their performance in the domain is self-relevant. On the contrary, less motivated people seem almost immune to the stereotype threat (Spencer et al., 1999; Stone et al., 1999). For instance, Whites who were threatened by the Asian math stereotype (Aronson et al. 1999) underperformed only if they were highly identified with math. People who endorse ingroup stereotypes, both implicitly (Nosek et al., 2009; Ramsey & Sekaquaptewa, 2011) and explicitly (Schmader et al., 2004), are more vulnerable to stereotype threat effects as well, perhaps because the stereotype remains chronically salient for them. Similarly, individuals who are high in stigma consciousness (i.e., the chronic self-awareness that one is a member of a stigmatized group) show larger performance decrements under threat than those who are low in stigma consciousness, because stigma consciousness leads targets to interpret more events in light of the stigmatized identity (Brown & Pinel 2003).

Other moderators of stereotype threat do not fit cleanly into the primary criteria: stereotype threat is more likely among individuals who are low self-monitors (Inzlicht et al., 2006), have an internal locus of control (Cadinu et al., 2006), lack a coping sense of humor (Ford et al., 2004), or have low working memory capacity (Régner et al., 2010). These variables may instead speak to processes underlying stereotype threat.

Processes underlying stereotype threat

To identify the processes that underlie stereotype threat effects, we should start by focusing on what mechanism is common among the tasks that stereotype threat affects. Although these tasks seem quite different, they share one important element: they are difficult, and therefore, they all require a certain degree of controlled attention, effortful processing, and self-regulation. Cognitive psychologists describe the mechanism that is responsible for this sort of efficient regulation as working memory (e.g., Engle, 2002). Research has established that stereotype threat experiences affect working memory. For example, Schmader and Johns (2003) found that manipulations of stereotype threat (e.g., describing a test as measuring quantitative or intellectual capacity) lower working memory capacity among stereotyped individuals (women and Latinos), while having no effect on those non-stereotyped (men and Whites). Furthermore, reductions in working memory capacity mediate the effects of the stereotype

threat manipulation on performance. In a similar vein, Beilock, Rydell, and McConnell (2007) showed that women under stereotype threat do more poorly on math problems but only if those problems are complex enough to require working memory resources.

Knowing that stereotype threat interferes with difficult tasks by consuming working memory suggests asking what processes are responsible for this effect. Researchers have proposed a variety of variables, including physiological responses, emotional reactions, cognitive and motivational processes. Studies suggest that each of these processes may contribute to impair situational working memory resources which are necessary for successful performance (Schmader, 2010; Schmader et al., 2008), even though, to date, none has received unequivocal support.

Physiological responses

Using skin conductance, skin temperature, blood pressure measures, and changes in heart rate, many studies have demonstrated that people under stereotype threat show stress-induced physiological arousal (e.g., Ben-Zeev et al., 2005; Croizet et al., 2004; Murphy et al., 2007; Osborne, 2007). According to the general stress and attention literature, physiological arousal has negative effects on performance for difficult tasks but positive effects for easy tasks (Zajonc, 1965). Indeed, when tasks are easy and do not require sustained attention provided by working memory, increased arousal elicited under stress can provide a boost in performance. However, as tasks become complex, stress-induced arousal has the potential to impair performance via its impact on working memory (e.g., Blair, 2006). These patterns parallel the finding that stereotype threat manipulations have their largest effects when tasks are difficult (Ben-Zeev et al., 2005; O'Brien & Crandall, 2003; Quinn & Spencer, 2001). Thus, physiological arousal appears to be an important mediator of stereotype threat effects.

Emotional reactions

Researchers have documented that the possibility to confirm an ingroup ability stereotype makes individuals anxious, and that heightened anxiety mediates, at least partially, the effects of stereotype threat on performance (Bosson et al., 2004; Brodish & Devine, 2009; Hoyt et al., 2010; Osborne, 2001; Spencer et al., 1999). Also attempts to control expressions of this anxiety may underlie stereotype threat processes (Johns et al.,

2008; Krendl et al., 2008; Wraga et al., 2007). Suppression of anxiety, indeed, depletes executive resources, contributing to lowered performance in the task at hand. Moreover, Keller and Dauenheimer (2003) have shown that girls' reports of frustration, disappointment, and sadness account for poor math performance under stereotype threat. In addition to producing anxiety, these negative emotions diminish the cognitive resources available that are necessary for maximal performance. Overall, this interpretation of the role of emotional reactions supports the working memory depletion account of stereotype threat.

Intrusive thoughts

Stereotype threat also activates intrusive thoughts such as thoughts of self-doubt (Steele & Aronson, 1995), negative expectancies (Stangor et al., 1998), and task-related worries (Beilock et al., 2007). Cadinu, Maass, Rosabianca, and Kiesner (2005) have found that women taking a difficult math test reported having more negative thoughts under stereotype threat. Moreover, the number of negative thoughts they had during the first half of the test mediated the effect of stereotype threat on performance during the second half of the test. It seems that intrusive thoughts distract (competing with the ongoing cognitive task), thus reducing the working memory capacity necessary to effectively meet the information-processing requirements of a task (Inzlicht et al., 2006). Moreover, as for emotional reactions, attempts to suppress intrusive thoughts during a performance situation (Johns et al., 2007; McGlone & Aronson, 2007; Spencer, 2003) are effortful and, again, may impair working memory (e.g., Muraven & Baumeister, 2000; Wenzlaff & Wegner, 2000).

Motivation

The possibility of giving a poor performance, thus confirming the in-group ability stereotype, may be particularly threatening to the self-concept. This threat to the self may lead targets to respond defensively, presumably in an effort to disconfirm the ability stereotype and/or to preserve a positive view of the self. For example, stereotype threat can produce a prevention focus (Higgins, 1998), a regulatory state in which individuals become vigilant to prevent failure. Under such conditions, people tend to use risk-averse means manifesting in higher performance accuracy and enhanced analytic thinking. People in a state of vigilance, however, tend to exhibit poorer performance on tasks that rely on creativity, openness,

flexibility, and speed (Seibt & Förster, 2004). Since many difficult tasks require both analytic thinking and a degree of openness and speed for successful completion, a prevention focus induced by stereotype threat can hinder performance. Other research suggests that individuals under threat are more likely to adopt performance-avoidance goals (Brodish & Devine, 2009), which lead to lowered absorption and interest in the task (Smith et al., 2007). Finally, a common defense for people who feel at risk of low performance is what psychologists refer to as self-handicapping. In an attempt to minimize the negative implications for the self of low performance, threatened individuals may attribute failure to external sources (e.g., prior stress; not enough sleep the night before) or make a task more difficult so as to provide a ready-made excuse for poor performance (Brown & Josephs 1999; Keller 2002; Steele & Aronson 1995).

Whereas these strategies momentarily harm performance, in the next section we will focus on potential long-term consequences of stereotype threat experiences, briefly discussing another way in which threatened people may cope with stereotype threat, namely disengagement from the stereotyped domain.

Potential long-term effects: the cyclical process of stereotype threat

Stereotype threat research has focused largely on the moderators and processes underlying the immediate impact of stereotypes on performance. Beyond the immediate impairment of the performance, however, stereotype threat has also been found to impair stereotyped students from building abilities in the first place (see Appel & Kronberger, 2012, for a review), by interfering with the encoding of material (Taylor & Walton, 2011), note taking and test preparation (Appel et al., 2011), the comprehension of rules (Rydell et al., 2010a), and the use of efficient strategies (Rydell et al., 2010b), thus contributing to lowering ability and setting the stage for later differences in educational attainment and advancement. Stereotype threat may also decrease the degree of engagement with a given domain. In an attempt to defend their self-esteem, individuals may temporarily disengage from a stereotyped domain, reducing the centrality of their self-concept in that domain (Major & Schmader, 1998; Major et al., 1998). There is empirical support for domain disengagement as an immediate consequence of stereotype threat. For example, women exposed to gender-stereotypic television commercials later avoid math

items in favor of verbal items on a standardized test, report less interest in the stereotyped domain, and fewer aspirations toward domain-related vocations (Davies et al. 2002, 2005; see also Murphy et al., 2007). Moreover, African Americans experiencing stereotype threat disengaged from academics in response to negative performance feedback (Nussbaum & Steele, 2007). It is worth noting that continuous disengagement resulting from repeated exposure to stereotype threatening situations may sometimes lead stereotyped individuals to disidentify entirely with a domain by continuously distancing themselves from their performance in that area.

Disidentification with areas of weakness is natural and adaptive. However, disidentification with academic domains has been hypothesized to be detrimental and maladaptive, potentially leading students to permanently opt out of that domain (Steele et al., 2002; Major et al. 1998; Major & Schmader, 1998; Nussbaum & Steele 2007; Schmader et al., 2001). Woodcock and collaborators (2012) have tested the stereotype threat-disidentification hypothesis across three academic years with a national longitudinal panel of undergraduate science students. Experience of stereotype threat has been found to be associated with scientific disidentification, which, in turn, predicted a significant decline in the intention to pursue a scientific career among Hispanic/Latino students (see also Beasley & Fischer, 2012, for similar results on women).

Other than contributing to group differences in educational advancement and participation, domain disidentification may have other detrimental consequences. Because stereotyped individuals leave domains where they are vulnerable to stereotype threat, they end up continuing to be underrepresented in these fields, thus perpetuating a lack of role models (a known buffer against threat effects; Marx & Roman, 2002; see also Maass et al., 2002). Underrepresentation in a given field may also contribute to a potentially unwelcoming environment for new entrants, in that simply being in the numerical minority in a stereotyped domain is stereotype threatening (Inzlicht & Ben-Zeev, 2000). Further, because stereotype threat operates largely below public awareness, its role in academic disparities remains hidden and underrepresentation may end to be attributed to some social groups' intrinsic lack of interest or ability. In this way, stereotype threat experiences may produce a cyclical long-term process, both promoting stereotypes that people belonging to some social groups have less ability, and contributing to the continuation of stereotype threat experiences, given that the lack of members of some social groups in some fields perpetuates ability stereotypes.

Conclusions

Stereotype threat research has come a long way since the original Steele and Aronson paper. Since then, we have learned much about the power of stereotype threat, observing its potential to affect various social groups in several important life domains. Our understanding of individual difference variables that render individuals more vulnerable to stereotype threat, as well as of the mechanisms underlying stereotype threat has also evolved.

One good news of this brief review is that the experience of stereotype threat is situational. The stereotype threat model, indeed, points out that members of social groups negatively stereotyped in a given area of competence can excel in that domain when the environment signals that the ability stereotype is irrelevant. In this sense, the stereotype threat model speaks not only about the problem but also about possible solutions. There is much that we can do to shape environments and reduce stereotype threat. To date, social psychologists have proposed many interventions. For example, task reframing interventions operate by changing the descriptions of tasks to minimize the relevance of a stereotype (e.g., Quinn & Spencer, 2001); threat cue removal interventions omit triggers known to activate negative stereotypes, such as moving demographic questions to the end of standardized tests to avoid priming negative stereotypes (e.g., Danaher & Crandall, 2008); role model interventions demonstrate that having more ingroup peers reduces stereotype threat and improves participation, aspirations, and persistence (Dasgupta, 2011; Dasgupta et al., 2015); self-affirmation interventions encourage people to focus on positive aspects of themselves in order to buffer against the threat, improving math performance and grade point averages over time (Cohen et al., 2006; Martens et al., 2006). Finally, mindset interventions encourage students to think about intelligence as something that can be increased, and this mindset results in greater enjoyment and value of education, and improved grades in school (Aronson et al., 2002; Good et al., 2003; see also Dweck, 2006). Overall, therefore, stereotype threat can be reduced with relatively small interventions.

Another good news is that the negative effects of stereotype threat can be significantly reduced by simply learning about stereotype threat research (Johns et al., 2005). For example, informing women that concern about stereotypes can influence concentration and thus performance on stereotype-relevant tests can improve performance, even in environments that continue to highlight stereotypes or induce testing concerns.

Finally, and perhaps most importantly, the more people know about the influence of stereotype threat, the less academic gaps may be attributed to deficits in domain-relevant capacities or interests.

References

Appel, M. & Kronberger, N. (2012). Stereotype threat and the achievement gap: Stereotype threat prior to test taking. Educational Psychology Review, 24, 609-635. doi: 10.1007/s10648-012-9200-4

Appel, M., Kronberger, N., & Aronson, J. (2011). Stereotype threat impairs ability building: Effects on test preparation among women in science and technology. European Journal of Social Psychology, 41, 904-913. doi:10.1002/ejsp.835

Aronson, J. (2002). Improving academic achievement: Impact of psychological factors on education. San Diego: Academic Press.

Aronson, J., Fried, C. B., & Good, C. (2002) Reducing the effects of stereotype threat on African American college students by shaping theories of intelligence. Journal of Experimental Social Psychology, 38,113-125.

Beasley, M. A., & Fischer, M. J. (2012). Why they leave: the impact of stereotype threat on the attrition of women and minorities from science, math and engineering majors. Social Psychology of Education, 15, 427-448.

Beilock, S. L., Jellison, W. A., Rydell, R. J., McConnell, A. R., & Carr, T. H. (2006). On the causal mechanisms of stereotype threat: Can skills that don't rely heavily on working memory still be threatened?. Personality and Social Psychology Bulletin, 32, 1059-1071. doi: 10.1177/0146167206288489 PMID: 16861310

Beilock, S. L., Rydell, R. J., & McConnell, A. R. (2007). Stereotype threat and working memory: Mechanisms, alleviation, and spillover. Journal of Experimental Psychology: General, 136, 256-76

Ben-Zeev, T., Fein, S., & Inzlicht, M. (2005). Arousal and stereotype threat. Journal of Experimental Social Psychology, 41, 174-81

Benbow, C. P., & Stanley, J. C. (1980). Sex differences in mathematical ability: Fact or artifact? Science, 210, 1262–1264.

Blair, C. (2006). How similar are fluid cognition and general intelligence? A developmental neuroscience perspective on fluid cognition as an aspect of human cognitive ability. Behavioral and Brain Sciences, 29, 109-160.

Bosson, J. K., Haymovitz, E. L., & Pinel, E. C. (2004). When saying and

doing diverge: The effect of stereotype threat on self-reported versus non-verbal anxiety. Journal of Experimental Social Psychology, 40, 247-255. doi: 10.1016/S0022-1031(03)00099-4

Brodish, A. B., & Devine, P. G. (2009). The role of performance-avoidance goals and worry in mediating the relationship between stereotype threat and performance. Journal of Experimental Social Psychology, 45, 180-185

Brown, R. P., & Josephs, R. A. (1999). A burden of proof: Stereotype relevance and gender differences in math performance. Journal of Personality and Social Psychology, 76, 246-257.

Brown, R. P., & Pinel, E. C. (2003). Stigma on my mind: Individual differences in the experience of stereotype threat. Journal of Experimental Social Psychology, 39, 626-633

Burgess, D. J., Warren, J., Phelan, S., Dovidio, J., & Van Ryn, M. (2010). Stereotype threat and health disparities: What medical educators and future physicians need to know. Journal of General Internal Medicine, 25, 169-177.

Cadinu, M., & Galdi, S. (2012): Gender differences in implicit gender self-categorization lead to stronger gender self-stereotyping by women than by men. European Journal of Social Psychology, 42, 546-552. doi: 10.1002/ejsp.1881

Cadinu, M., Maass, A., Frigerio, S., Impagliazzo, L., & Latinotti, S. (2003). Stereotype threat: The effect of expectancy on performance. European Journal of Social Psychology, 33, 267-285.

Cadinu, M., Maass, A., Lombardo, M., & Frigerio, S. (2006). Stereotype threat: The moderating role of locus of control beliefs. European Journal of Social Psychology, 36, 183–197.

Cadinu, M., Maass, A., Rosabianca, A., & Kiesner, J. (2005). Why do women underperform under stereotype threat? Psychological Science, 16, 572-578.

Carr, P. B., & Steele, C. M. (2010). Stereotype threat affects financial decision making. Psychological Science, 21, 1411-1416.

Cohen, G. L., Garcia, J., Apfel, N., & Master, A. (2006). Reducing the racial achievement gap: a social psychological intervention. Science, 313, 1307-1310.

Croizet, J. C., & Claire, T. (1998). Extending the concept of stereotype threat to social class: The intellectual underperformance of students from low socioeconomic backgrounds. Personality and Social Psychology Bulletin, 24, 588–594.

Croizet, J. C., Désprés, G., Gauzins, M. E., Huguet, P., Leyens, J. P., & Méot,

A. (2004). Stereotype threat undermines intellectual performance by triggering a disruptive mental load. Personality and Social Psychology Bulletin, 30, 721-731

Croizet, J. C., & Millet, M. (2012). Social class and test performance: From stereotype threat to symbolic violence and vice versa. In: M. Inzlicht, & T. Schmader, (eds) Stereotype threat: Theory, process, and application. New York: Oxford University Press

Danaher, K., & Crandall, C. S. (2008). Stereotype threat in applied settings re-examined. Journal of Applied Social Psychology, 38,1639-1655.

Dasgupta, N. (2011). Ingroup experts and peers as social vaccines who inoculate the self-concept: the stereotypes inoculation model. Psychological Inquiry, 22, 231-246.

Dasgupta, N., Scircle, M. M., & Hunsinger, M. (2015). Female peers in small work groups enhance women's motivation, verbal participation, and career aspirations in engineering. PNAS, 112, 4498-4933.

Davies, P. G., Spencer, S. J., Quinn, D. M., & Gerhardstein, R. (2002). Consuming images: how television commercials that elicit stereotype threat can restrain women academically and professionally. Personality and Social Psychology Bulletin, 28, 1615-1628

Davies, P. G., Spencer, S. J., & Steele, C. M. (2005). Clearing the air: Identity safety moderates the effects of stereotype threat on women's leadership aspirations. Journal of Personality and Social Psychology, 88, 276-287

Dweck, C. (2006). Mindset the new psychology of success. Random House.

Engle, R. W. (2002). Working memory capacity as executive attention. Current Directions in Psychological Science, 11, 19-23

Ford, T. E., Ferguson, M. A., Brooks, J. L., & Hagadone, K. M. (2004). Coping sense of humor reduces effects of stereotype threat on women's math performance. Personality and Social Psychology Bulletin, 30, 643-653

Galdi, S., Cadinu, M., & Tomasetto, C. (2014): The roots of stereotype threat: When automatic associations disrupt girls' math performance. Child Development, 85, 250-263. doi:10.1111/cdev.12128

Good, C., Aronson, J., & Inzlicht, M. (2003). Improving adolescents' standardized test performance: an intervention to reduce the effects of stereotype threat. Journal of Applied Developmental Psychology, 24, 645-662.

Gupta, V. K., Goktan, A. B., & Gunay, G. (2014). Gender differences in evaluation of new business opportunity: A stereotype threat perspective. Journal of Business Venturing, 29, 273-288. doi:10.1016/j.jbusvent.2013.02.002

Herrnstein, R. J., & Murray, C. (1994). The bell curve. Intelligence and class structure in American life. New York: Free Press.

Higgins, E. T. (1998). Promotion and prevention: Regulatory focus as a motivational principle. In M. P. Zanna (Ed.), Advances in experimental social psychology (Vol. 30, pp. 1-46). San Diego, CA: Academic Press.

Hoyt, C. L., Johnson, S. K., Murphy, S. E., & Skinnell, K. H. (2010). The impact of blatant stereotype activation and group sex-composition on female leaders. Leadership Quarterly, 21, 716-732.

Huguet, P., & Régner, I. (2009). Counter-stereotypic beliefs in math do not protect school girls from stereotype threat. Journal of Experimental Social Psychology, 45, 1024-1027

Inzlicht, M., & Ben-Zeev, T. (2000). A threatening intellectual environment: Why females are susceptible to experiencing problem-solving deficits in the presence of males. Psychological Science, 11, 365-371

Inzlicht, M., & Ben-Zeev, T. (2003). Do high-achieving female students underperform in private? The implications of threatening environments on intellectual processing. Journal of Educational Psychology, 95, 796-805.

Inzlicht, M., & Good, C. (2006). How environments can threaten academic performance, self-knowledge, and sense of belonging. In Stigma and group inequality (pp. 143-164). Psychology Press.

Inzlicht, M., McKay, L., & Aronson, J. (2006). Stigma as ego-depletion: How being the target of affects self-control. Psychological Science, 17, 262–269.

Johns, M., Inzlicht, M., & Schmader, T. (2008). Stereotype threat and executive resource depletion: examining the influence of emotion regulation. Journal of Experimental Psychology: General, 137, 691-705

Johns, M., Schmader, T., & Martens, A. (2005). Knowing is half the battle: teaching stereotype threat as a means of improving women's math performance. Psychological Science 16,175-179

Keller, J. (2002). Blatant stereotype threat and women's math performance: self-handicapping as a strategic means to cope with obtrusive negative performance expectations. Sex Roles 47, 193-198

Keller, J, & Dauenheimer, D. (2003). Stereotype threat in the classroom: Dejection mediates the disrupting threat effect on women's math performance. Personality and Social Psychology Bulletin, 29, 371-381

Kray, L. J., Galinsky, A. D., & Thompson, L. (2002). Reversing the gender gap in negotiations: An exploration of stereotype regeneration. Organizational Behavior and Human Decision Processes, 87, 386-409

Krendl, A. C., Richeson, J. A., Kelley, W. M., & Heatherton, T. F. (2008). The negative consequences of threat: A functional magnetic resonance imaging investigation of the neural mechanisms underlying women's underperformance in math. Psychological Science, 19, 168-175.

Levy, B. (1996). Improving memory in old age through implicit self-stereotyping. Journal of Personality and Social Psychology, 71, 1092-1107.

Logel, C., Walton, G. M., Spencer, S. J., Iserman, E. C., von Hippel, W., & Bell, A. E. (2009). Interacting with sexist men triggers social identity threat among female engineers. Journal of Personality and Social Psychology, 96, 1089-1103

Maass, A., & Cadinu, M. (2003). Stereotype threat: When minority members underperform. European Review of Social Psychology, 14, 243-275.

Maass, A., Cadinu, M., Verga, S., & Crimaldi, S. (2002). Preventing stereotype-induced memory deficits among the elderly: The Levi-Montalcini effect (unpublished manuscript).

Major, B., & Schmader, T. (1998). Coping with stigma through psychological disengagement. In J. Swim & C. Stangor (Eds.), Stigma: The target's perspective. New York: Academic Press.

Major, B., Spencer, S., Schmader, T.,Wolf, C., & Crocker, J. (1997). Coping with negative stereotypes about intellectual performances: The role of psychological disengagement. Personality and Social Psychology Bulletin, 24, 34–50.

Major, B., Spencer, S. J., Schmader, T., Wolfe, C. T., & Crocker, J. (1998). Coping with negative stereotypes about intellectual performance: The role of psychological disengagement. Personality and Social Psychology Bulletin, 24, 34-50.

Martens, A., Johns, M., Greenberg, J., & Schimel, J. (2006). Combating stereotype threat: the effect of self-affirmation on women's intellectual performance. Journal of Experimental Social Psychology, 42, 236-243.

Marx, T., & Roman, J. S. (2002). Female role models: Protecting women's math test performance. Personality and Social Psychology Bulletin, 28, 1183-1193.

Mazerolle, M., Régner, I., Morisset, P., Rigalleau, F., & Huguet, P. (2012). Stereotype threat strengthens automatic recall and undermines controlled processes in older adults. Psychological Science, 23, 723727

McGlone, M. S., & Aronson, J. (2007). Forewarning and forearming stereotype-threatened students. Communication Education, 56, 119-133.

Muraven, M., & Baumeister, R. F. (2000). Self-regulation and depletion of limited resources: Does self-control resemble a muscle? Psychological Bulletin, 126, 247-259.

Murphy, M. C., Steele, C. M., &Gross, J. J. (2007). Signaling threat: how situational cues affect women in math, science, and engineering settings. Psychological Science, 18, 879-885

Nadler, J. T., & Clark, M. H. (2011). Stereotype threat: A meta-analysis comparing African Americans to Hispanic Americans. Journal of Applied Social Psychology, 41, 872-890.

Nguyen, H. H. D., & Ryan, A. M. (2008). Does stereotype threat affect test performance of minorities and women? A meta-analysis of experimental evidence. Journal of Applied Psychology, 93, 1314-1334

Nosek, B. A., Smyth, F. L., Sriram, N., Lindner, N. M., Devos, T., Ayala, A., … & Greenwald, A. G. (2009). National differences in gender-science stereotypes predict national sex differences in science and math achievement. PNAS, 106, 10593-10597.

Nussbaum, A. D., & Steele, C. M. (2007). Situational disengagement and persistence in the face of adversity. Journal of Experimental Social Psychology, 43, 127-134

O'Brien, L. T., & Crandall, C. S. (2003). Stereotype threat and arousal: effects on women's math performance. Personality and Social Psychology Bulletin, 29, 782-788

Osborne, J. W. (2001). Testing stereotype threat: Does anxiety explain race and sex differences in achievement?. Contemporary Educational Psychology, 26, 291-310.

Osborne, J. W. (2007). Linking stereotype threat and anxiety. Educational Psychology, 27, 135-154.

Purdie-Vaughns, V., Steele, C. M., Davies, P. G., Ditlmann, R., & Crosby, J. R. (2008). Social identity contingencies: How diversity cues signal threat or safety for African Americans in mainstream institutions. Journal of Personality and Social Psychology, 94, 615-630

Quinn, D. M., & Spencer, S. J. (2001). The interference of stereotype threat with women's generation of mathematical problem-solving strategies. Journal of Social Issues, 57, 55-71.

Ramsey, L. R., & Sekaquaptewa, D. (2011). Changing stereotypes, changing grades: A longitudinal study of stereotyping during a college math course. Social Psychology of Education, 14, 377-387.

Régner, I., Smeding, A., Gimmig, D., Thinus-Blanc, C., Monteil, J. M., & Huguet, P. (2010). Individual differences in working memory moderate stereotype-threat effects. Psychological Science, 21, 1646-1648

Rydell, R. J., Rydell, M. T., & Boucher, K. L. (2010a). The effect of negative performance stereotypes on learning. Journal of Personality and Social Psychology, 99, 883-896.

Rydell, R. J., Shiffrin, R. M., Boucher, K. L., Van Loo, K., & Rydell, M. T. (2010b). Stereotype threat prevents perceptual learning. PNAS, 107, 14042-14047.

Scafidi, T., & Bui, K. (2010). Gender similarities in math performance from middle school through high school. Journal of Instructional Psychology, 37, 252-255.

Schmader, T. (2002). Gender identification moderates stereotype threat effects on women's math performance. Journal of Experimental Social Psychology, 38, 194-201.

Schmader, T. (2010). Stereotype threat deconstructed. Current Directions in Psychology, 19, 14-18

Schmader, T., & Johns, M. (2003). Converging evidence that stereotype threat reduces working memory capacity. Journal of Personality and Social Psychology, 85, 440-452.

Schmader, T., Major, B., & Gramzow, R. H. (2001). Coping with ethnic stereotypes in the academic domain: Perceived injustice and psychological disengagement. Journal of Social Issues, 57, 93-111.

Schmader, T., Johns, M., & Barquissau, M. (2004). The costs of accepting gender differences: The role of stereotype endorsement in women's experience in the math domain. Sex Roles, 50, 835-850. doi: 10.1023/B:SERS.0000029101.74557.a0

Schmader, T., Johns, M., & Forbes, C. (2008). An integrated process model of stereotype threat effects on performance. Psychological Review, 115, 336-356

Seibt, B., & Forster, J. (2004). Stereotype threat and performance: How self-stereotypes influence processing by inducing regulatory foci. Journal of Personality and Social Psychology, 87, 38-56

Shapiro, J. R. (2011). Different groups, different threats: a multi-threat approach to the experience of stereotype threats. Personality and Social Psychology Bulletin, 37, 464-480

Shapiro, J. R, & Neuberg, S. L. (2007). From stereotype threat to stereotype threats: implications of a multi-threat framework for causes, moderators, mediators, consequences, and interventions. Personality and Social Psychology Review, 11, 107-130

Shih, M., Pittinsky, T. L., & Ambady, N. (1999). Stereotype susceptibility: identity salience and shifts in quantitative performance. Psychological Science, 10, 80-83.

Skorich, D. P., Webb, H., Stewart, L., Kostyanaya, M., Cruwyz, T., McNeill, K., et al. (2013). Stereotype threat and hazard perception among provisional license drivers. Accident Analysis & Prevention, 54, 39-45. doi: 10.1016/j.aap.2013.02.002 PMID: 23474236

Smith, J. L., Sansone, C., & White, P. H. (2007). The stereotyped task engagement process: The role of interest and achievement motivation. Journal of Educational Psychology, 99, 99-114.

doi:10.1037/0022-0663.99.1.99

Spencer, S. J., Steele, C. M., & Quinn, D. (1999). Stereotype threat and women's math performance. Journal of Experimental Social Psychology, 35, 4-28.

Stangor, C., Carr, C., & Kiang, L. (1998). Activating stereotypes undermines task performance expectations. Journal of Personality and Social Psychology, 75, 1191-1187.

Steele, C. M. (1997). A threat in the air: How stereotypes shape intellectual identity and performance. American Psychologist, 52, 613-629.

Steele, C. M., & Aronson, J. (1995). Stereotype threat and the intellectual test performance of African Americans. Journal of Personality and Social Psychology, 69, 797-811.

Steele, C. M., Spencer, S. J., & Aronson, J. (2002a). Contending with group image: the psychology of stereotype and social identity threat. Advances in Experimental Social Psychology, 34, 379-440

Steele, J., James, J. B., Barnett, R. C. (2002b). Learning in a man's world: Examining the perceptions of undergraduate women in male-dominated academic areas. Psychology of Women Quarterly, 26, 46-50

Stone, J. (2002). Battling doubt by avoiding practice: The effects of stereotype threat on self-handicapping in white athletes. Personality and Social Psychology Bulletin, 28, 1667-1678.

Stone, J., Lynch, C. I., Sjomeling, M., & Darley, J. M. (1999). Stereotype threat effects on Black and White athletic performance. Journal of Personality and Social Psychology, 77, 1213-1227. doi: 10.1037/0022-3514.77.6.1213

Summers, L. (2005, January 14). Remarks at NBER conference on diversifying the science and engineering workforce. Retrieved from http://www.harvard.edu/president/speeches/summers_2005/nber.php

Tajfel, H., & Turner, J. C. (1986). The social identity theory of inter-group behavior. In S. Worchel & L. W. Austin (Eds.), Psychology of intergroup relations. Chicago: Nelson-Hall.

Taylor, V. J., & Walton, G. M. (2011). Stereotype threat undermines academic learning. Personality and Social Psychology Bulletin, 37, 1055-1067. doi: 10.1177/0146167211406506

Tenenbaum, H. R., & Leaper, C. (2003). Parent-child conversations about science: The socialization of gender inequities?. Developmental Psychology, 39, 34.

Wenzlaff, R. M., & Wegner, D. M. (2000) Thought suppression. Annual Review of Psychology, 51, 59-91.

Wheeler, S. C., & Petty, R. E. (2001). The effects of stereotype activation on behaviour: A review of possible mechanisms. Psychological Bulletin, 127, 797-826

Woodcock, A., Hernandez, P. R., Estrada, M., & Schultz, P. (2012). The consequences of chronic stereotype threat: Domain disidentification and abandonment. Journal of Personality and Social Psychology, 103, 635-646

Wout, D., Danso, H., Jackson, J., & Spencer, S. (2008). The many faces of stereotype threat: Group- and self-threat. Journal of Experimental Social Psychology, 44, 792-799.

Wout, D., Shih, M. J., Jackson, J. S., & Sellers, R. M. (2009). Targets as perceivers: How people determine when they will be negatively stereotyped. Journal of Personality and Social Psychology, 96, 349-362

Wraga, M., Helt, M., Jacobs, E., & Sullivan, K. (2007). Neural basis of stereotype-induced shifts in women's mental rotation performance. Social Cognitive and Affective Neuroscience, 2, 12-19.

Yeung, N. C. J., & von Hippel, C. (2008). Stereotype threat increases the likelihood that female drivers in a simulator run over jaywalkers. Accident Analysis & Prevention, 40, 667-674. doi: 10.1016/j.aap.2007.09.003

Zajonc, R. B. (1965). Social facilitation. Science, 149, 269-274

21. Beyond their semantic and evaluative tone: Derogatory group labels as a social tool of disempowerment

Andrea Carnaghi and Mauro Bianchi
University of Trieste, Italy

Derogatory group labels (i.e., DGLs) frame group memberships (e.g., a Jew) in an offensive fashion (e.g., a kike), thus allowing an individual or a group to be insulted by disparaging their social identity (Carnaghi & Bianchi, 2017; O'Dea et al., 2015).

In this chapter, we will review three distinct, though related, theoretical perspectives that have conceptualized DGLs differently. The first perspective is focused on the lexical aspects of DGLs, and suggests that label semantics account for both the appraisal and the psychological consequences of exposure to DGLs (Mullen & Leader, 2005). The second perspective stems from Allport's insight suggesting that DGLs are 'emotionally toned labels' (Allport, 1954/1979, p.181). Accordingly, DGLs are framed as 'a telegraphic shorthand for hostile prejudice' (Mullen & Leader, 2005), thus highlighting the close interplay between DGLs and prejudice. The third perspective was introduced by Maass and later developed by other scholars; this perspective has emphasized the role of DGLs in highlighting the lower status of a stigmatized group and the disempowering of that group (Carnaghi & Maass, 2007a; Wang et al., 2017).

Rather than being mutually exclusive, these perspectives have focused on specific facets of DGLs and have highlighted a panoply of unique psychological consequences associated with overhearing such labels.

The semantics of DGLs

The mispronunciation of ethnic group labels (e.g., Paki for Pakistani), as well as the intentional phonological alteration of the first names of members of an ethnic group, and the conceptualization of an ethnic group with the most frequent proper name (e.g., Giuseppe for Italian people) appear to be the most common ways to denigrate and engage in verbal (micro)aggressions towards ethnic groups (Allen, 1988; Kholi & Solórzano, 2012).

A more blatant fashion to verbally denigrate outgroups is to replace the category labels which are merely meant to describe the groups (e.g., proper, preferred labels of a group, such as 'woman') with 'an altogether different name with negative semantics' (Allen, 1988; p. 217). The target group is conflated with the name of a famous, and possibly negative, group member (e.g., Adolf for German people), or called by food names (e.g., baguette for French people) and animal metaphors (e.g., frogs for French people).

Animal metaphors have been regarded as taboo words, and their use has been long sanctioned or limited in various cultures because of the acknowledged harm that could result from their use (Jay, 2009). Examples of animal metaphors as DGLs can be easily detected in the history of genocides and violence: 'Jews' were depicted as rats, snakes, and parasites in 'La difesa della Razza' -a fascist publication issued in Italy in 1938- (Volpato, 2013); Tutsis were referred to as 'cockroaches' during the conflict in Rwanda (Bell-Fialkoff, 1996).

Mullen and Leader (2005, p. 195) noticed that animal metaphors are disproportionally used to name the outgroups compared to the ingroup, suggesting the risk run by the outgroup of being dehumanized: 'ingroup gives itself [names] often derive from native words for 'real humans' or 'the people', whereas the ethnophaulisms [i.e., DGLs] given to outgroups are often derived from native words for 'beast' and 'animal' (Fried, 1975)'. Along this line, Haslam and colleagues (2011) suggested that the perceived offensiveness of DGLs issued by animal metaphors is accounted for by the unique meaning conveyed by the various metaphors. In particular, the offensiveness of animal metaphors used as DGLs is strongly entrenched both in their content of depravity and unpleasantness and in the dehumanizing meaning they suggest. Indeed, some animal metaphors, such as those comparing humans with rats and snakes, convey (moral) disgust. In addition, animal metaphors that are judged highly offensive

are those that succeed in turning humans into animal-like beings, as in the case of metaphors indicating stupidity (e.g., "monkey").

The key role of the semantics of DGLs in enacting and perpetuating social exclusion has been established by several archives and laboratory studies carried out by Mullen (Mullen & Johnson, 1993; 1995; Mullen & Leader, 2005; Mullen et al., 2000; Mullen, 2001). According to Mullen and colleagues, DGLs are organized along two basic dimensions. The former deals with the semantics of DGLs and is referred to as complexity, and the latter points to the evaluative tone of DGLs and is termed valence. Specifically, DGLs for some ethnic groups span several semantic clusters. For example, Italian immigrants in the United States are referred to by personal names (e.g., Giuseppe, Maria), by the names of typical Italian food (e.g., pasta, spaghetti), by the names of criminal associations (e.g., mafia), signaling a relatively higher level of complexity in cognitive representation. For other ethnic groups, DGLs are grouped around a few semantic clusters, implying a relatively lower level of complexity in cognitive representation (e.g., Belgian immigrants in the United States are addressed by some semblance of their group name, "belgeek" and "flamingo" – Mullen & Leader, 2005). Also, for the same group, some semantic clusters are more negative (e.g., mafia) than others (e.g., spaghetti). The overall evaluation given to the DGLs of a group is derived from the average of the valence associated with each cluster, so that some groups are represented more negatively than others.

Archival research carried out by Mullen and Smyth (2004) and Mullen and Rice (2003) has demonstrated that the complexity of the DGLs, more than the valence of these DGLs, better predicted the social exclusion of ethnic groups across 150 years. This pattern of results was later corroborated by experimental research. Mullen, Leader, and Rice (2004) orthogonally manipulated the complexity and valence of DGLs pointing to a group. Results indicated that decreasing the complexity of the DGLs which participants used to call a group, more so than increasing the negativity of such DGLs, promoted the stronger exclusion of that group.

Overall, this line of research warns of the crucial role played by the semantics of DGLs in their perceived offensiveness, as well as of the complexity of DGL semantics in perpetrating social exclusion.

The evaluative tone of DGLs

According to Allport, when DGLs are used 'we can be almost certain that the speaker intends not only to characterize the person's memberships but also to disparage and reject him' (p.181, italics in original; Allport, 1954/1979).

This conceptualization puts forward three fundamental aspects of the DGLs. First, while both DGLs and category labels point to the group membership of their targets, DGLs convey a harsher evaluative tone than category labels (Carnaghi & Maass, 2007a; 2007b). Second, while common insult convey a negative tone that targets an individual identity, DGLs disparage the targets' social identity (Bianchi et al., 2019; O'Dea et al., 2017). Third, DGLs can be linguistic discriminatory tools that bear prejudiced meanings and can reinforce prejudiced attitudes (Carnaghi & Maass, 2007a).

Carnaghi and Maass (2007b) proposed that as DGLs and category labels point to the same group membership, they should activate similar stereotypical content. However, due to their prejudiced undertones, DGLs should activate less flattering evaluative reactions than category labels. To test this idea, Carnaghi and Maass (2007b) analyzed the manner in which heterosexual participants appraised category and DGLs referring to gay men. Specifically, the authors relied on a lexical decision task in which heterosexual participants were shown a screen in which words (i.e., targets) or strings of letters which did not make up a real word. Participants were asked to quickly decide whether a word was real or not by pressing appropriate keys on a keyboard. The targets varied orthogonally in their level of stereotypically to the category of gay men (i.e., stereotypical targets, such as 'effeminate'; counter-stereotypical targets, such as 'virile', and irrelevant targets, such as 'greedy') and valence (i.e., positive and negative targets). Each target was preceded (i.e., primed) once by a DGL (e.g., 'fag') and once by a category label (e.g., 'gay'). Participants were faster in recognizing stereotypical targets as words regardless of the prime. This implies that when primed with category labels or DGLs referring to gay men, the stereotypes were equally activated. Importantly, DGLs in comparison to category labels inhibited the accessibility of positive contents. Corroborating these results, albeit by using a self-report measure, Bianchi and colleagues (2019) found that both DGLs and category were similar in pointing to the group membership of their targets, but DGLs were appraised as more offensive and their use as less socially acceptable than category labels.

A key test of this perspective is the comparison between DGLs and common insults. Both types of labels slander their targets, but only DGLs point to a specific social group while common insults are not tied nor address any specific group. With regard to this, O'Dea and colleagues (O'Dea et al., 2015; O'Dea et al., 2017) presented participants with vignettes describing an interracial interaction in which a White man referred to a Black man using either a non-racial term (e.g., 'buddy'), a common insult (e.g., 'asshole'), or a racial DGL (e.g., 'nigger'). Results showed that participants' ratings of offensiveness would increase from non-racial terms to common insults, peaking with DGLs. In a similar vein, Bianchi and colleagues (2019) found that DGLs were appraised as more offensive and less socially acceptable than common insults.

A crucial claim of this theoretical perspective is that only DGLs, but not category labels or common insults, 'precipitate subsequent hostile prejudice, exclusion and violence towards the target outgroup' (Mullen & Leader, 2005; p. 194). In support of this claim, Fasoli and colleagues (2016) showed that heterosexual participants increased their physical distance from a gay man after being exposed to a DGL (e.g., 'faggot'), compared with a category label (e.g., 'gay') and a common insult (e.g., 'asshole'). Hence, DGLs appear to increase participants' prejudice-based responses, as physical distance represents an unobtrusive measure of participants' prejudice. Bianchi and colleagues (2019) tested the reverse relationship, namely whether enhanced levels of prejudice towards specific groups affected the appraisal of DGLs referring to those groups. Bianchi and colleagues (2019) found that participants' level of prejudice reduced the perceived offensiveness of DGLs, and in turn enhanced their social acceptability. Importantly, participants' prejudice toward the groups that were targets of DGLs was unrelated to both the offensiveness and social acceptability of common insults.

Overall, these results point to the specificity of DGLs as linguistic devices that carry a negative evaluative tone and perpetuate prejudice and discrimination.

DGLs as a disempowering social device

DGLs are words typically created by majority members to name minorities. This genesis of DGLs gives rise to two constitutive features of DGLs. First, DGLs strengthen intergroup contexts, leading to a cascade of cognitive and social biases typically observed when intergroup relations

are worsened (Carnaghi & Maass, 2007a). Second, DGLs are words that highlight the inferiority and stigmatized status of the outgroup and cause the victim to feel disempowered (Henry et al., 2014; Wang et al., 2017).

Carnaghi & Maass (2007a) relied on a lexical decision task in which participants were presented with some words (i.e., targets) and some strings of letters that did not form a real word. Importantly, targets comprised words that pointed to inter-group context (e.g., cooperation and antagonism), and words that were pretested as unrelated to the intergroup (e.g., holiday and vomit). Each target was preceded once by a category label (e.g., 'gay'), and once by a DGL (e.g., 'fag'). Analyses on reaction times revealed that no difference was observed between a DGL and a category label for words that were unrelated to the intergroup relations, while reaction times tended to be faster for words related to intergroup relations when following a DGL rather than a category label. Therefore, exposure to DGLs, rather than category labels, is highly likely to frame the social context in an intergroup fashion.

The idea that DGLs are effective in stressing the inter-group nature of social interactions has been corroborated by studies showing that DGLs enhance ingroup favoritism (Fasoli, Maass, & Carnaghi, 2015). Indeed, Fasoli and colleagues exposed heterosexual participants either to a DGL or a category label (i.e., homosexual) addressing gay men, and asked them to allocate fictitious funds to two prevention programs. One program was stereotypically associated with heterosexuals, namely a sterility prevention program, while the other was stereotypically associated with gay men, that is an AIDS-HIV prevention program. Findings suggested stronger ingroup favoritism, namely a preference to allocate funds to prevention programs for heterosexual rather than gay individuals, after the exposure to DGLs than to category labels.

Subsequent studies have highlighted the power of DGLs not only in accentuating the intergroup nature of social interactions, but also in worsening the quality of those interactions. Indeed, DGLs are not only effective in increasing favoritism toward the ingroup (Fasoli et al., 2015), but are particularly effective in promoting disdain toward the outgroup (Bilewicz & Soral, 2020). It is worth noting that while ingroup preference is, at least to varying degrees, a pervasive side effect of intergroup contexts per se, outgroup hatred is more likely to emerge in deteriorated group contexts, such as those characterized by conflict and enmity (Mummendey & Otten, 1998; Weisel & Böhm, 2015).

The potential for DGLs to emphasize and degrade the nature of intergroup relations is likely due to their potential to highlight the minority and stigmatized status of the group they target. Indirect evidence to support this conjecture are findings attesting that the perceived offensiveness of DGLs depends on the status of the group they target (Henry et al., 2014), and on their ability to disempower the targeted minority (Wang et al., 2017). Hence, it is most plausible that DGLs function as signals of the differential status of groups. Such accentuation of status asymmetry likely leads members of the dominant group to discriminate against (outgroup disdain) members of the minority group, as it occurs in a clearly asymmetrical group status context (Sachdev & Bourhis, 1987).

Conclusion

We briefly illustrated three different non-mutually exclusive perspectives in the study of DGLs.

The first perspective takes into consideration the semantic content conveyed by DGLs and how this content differs from the one conveyed by category labels. Significant to this perspective are studies that focus on the lexical aspect of DGLs and show how the contents and the semantic complexity of the DGLs shape bystanders' reactions to such labels.

The second perspective considers DGLs to be discriminatory devices that express negative evaluative tones and perpetuate prejudice. Studies from this perspective have focused on the comparison between DGLs and category and/or common insults and show a differential appraisal of DGLs as conveyors of negative meanings and reinforcers of prejudice.

The third perspective takes into account the complex intergroup dynamics that embed the use of DGLs and their consequences for status and power. In this perspective, DGLs are mainly products of the dominant majority used in the maintenance of status hierarchies via the exacerbation of intergroup dynamics and the exclusion of comparatively less powerful minorities.

References

Allen, J. (1988). Natural language understanding. Benjamin-Cummings Publishing Co., Inc.

Allport, G. W. (1954/1979). The nature of prejudice. Reading, MA: Perseus Books.

Bell-Fialkoff, A. (1996). Ethnic cleansing. New York, NY: St. Martins Press.

Bianchi, M., Carnaghi, A., Piccoli, V., Stragà, M., & Zotti, D. (2019). On the descriptive and expressive function of derogatory group labels: An experimental test. Journal of Language and Social Psychology, 38(5-6), 756-772. https://doi.org/10.1177/0261927X19867739

Bilewicz, M., & Soral, W. (2020). Hate speech epidemic. The dynamic effects of derogatory language on intergroup relations and political radicalization. Political Psychology, 41, 3-33. https://doi.org/10.1111/pops.12670

Carnaghi, A., & Bianchi, M. (2017). Derogatory group labeling. In Oxford Research Encyclopedia of Communication. https://doi.org/10.1093/acrefore/9780190228613.013.435

Carnaghi, A., & Maass, A. (2007a). Derogatory language in intergroup context: Are "gay" and "fag" synonymous?. In Y. Kashima, K. Fiedler, & P. Freytag (Eds.), Stereotype Dynamics (pp. 126-143). Psychology Press.

Carnaghi, A., & Maass, A. (2007b). In-group and out-group perspectives in the use of derogatory group labels: Gay versus fag. Journal of Language and social Psychology, 26(2), 142-156. https://doi.org/10.1177/0261927X07300077

Jay, T. (2009). The utility and ubiquity of taboo words. Perspectives on Psychological Science, 4, 153-161. https://doi.org/10.1111/j.1745-6924.2009.01115.x

Fasoli, F., Maass, A., & Carnaghi, A. (2015). Labelling and discrimination: Do homophobic epithets undermine fair distribution of resources?. British Journal of Social Psychology, 54(2), 383-393. https://doi.org/10.1111/bjso.12090

Fasoli, F., Paladino, M. P., Carnaghi, A., Jetten, J., Bastian, B., & Bain, P. G. (2016). Not "just words": Exposure to homophobic epithets leads to dehumanizing and physical distancing from gay men. European Journal of Social Psychology, 46(2), 237-248. https://doi. org/10.1002/ejsp.2148

Fried, M. (1975). The notion of tribe. Menlo Park, CA: Cummings.

Haslam, N., Loughnan, S., & Sun, P. (2011). Beastly: What makes animal metaphors offensive?. Journal of Language and Social Psychology, 30(3), 311-325. https://doi.org/10.1177/0261927X11407168

Henry, P. J., Butler, S. E., & Brandt, M. J. (2014). The influence of target group status on the perception of the offensiveness of group-based slurs. Journal of Experimental Social Psychology, 53, 185-192. https://doi.org/10.1016/j.jesp.2014.03.01

Mummendey, A., & Otten, S. (1998). Positive–negative asymmetry in social discrimination. European Review of Social Psychology, 9(1), 107-143. https://doi.org/10.1080/14792779843000063

Kohli, R., & Solórzano, D. G. (2012). Teachers, please learn our names!: Racial microagressions and the K-12 classroom. Race Ethnicity and Education, 15(4), 441-462. https://doi.org/10.1080/13613324.2012.674026

Mullen, B. (2001). Ethnophaulisms for ethnic immigrant groups. Journal of Social Issues, 57(3), 457-475. https://doi.org/10.1111/0022-4537.00223

Mullen, B., & Johnson, C. (1993). Cognitive representation in ethnophaulisms as a function of group size: The phenomenology of being in a group. Personality and Social Psychology Bulletin, 19(3), 296-304. https://doi.org/10.1177/0146167293193006

Mullen, B., & Johnson, C. (1995). Cognitive representation in ethnophaulisms and illusory correlation in stereotyping. Personality and Social Psychology Bulletin, 21(5), 420-433. https://doi.org/10.1177/0146167295215001

Mullen, B., & Leader, T. (2005). Linguistic Factors: Antilocutions, Ethnonyms, Ethnophaulisms, and Other Varieties of Hate Speech. In J. F. Dovidio, P. Glick, & L. A. Rudman (Eds.), On the nature of prejudice: Fifty years after Allport (pp. 192–207). Blackwell Publishing. https://doi.org/10.1002/9780470773963.ch12

Mullen, B., & Rice, D. R. (2003). Ethnophaulisms and exclusion: The behavioral consequences of cognitive representation of ethnic immigrant groups. Personality and Social Psychology Bulletin, 29(8), 1056-1067. https://doi.org/10.1177/0146167203254505

Mullen, B., Rozell, D., & Johnson, C. (2000). Ethnophaulisms for Ethnic Immigrant Groups: Cognitive Representation of the Minority' and the Foreigner'. Group Processes & Intergroup Relations, 3(1), 5-24. https://doi.org/10.1177/1368430200031001

Mullen, B., & Smyth, J. M. (2004). Immigrant suicide rates as a function of ethnophaulisms: Hate speech predicts death. Psychosomatic Medicine, 66(3), 343-348. https://doi.org/10.1097/01.psy.0000126197.59447.b3

O'Dea, C. J., Miller, S. S., Andres, E. B., Ray, M. H., Till, D. F., & Saucier, D. A. (2015). Out of bounds: Factors affecting the perceived offensiveness of racial slurs. Language Sciences, 52, 155-164. https://doi.org/10.1016/j.langsci.2014.09.005

O'Dea, C. J., & Saucier, D. A. (2017). Negative emotions versus target descriptions: Examining perceptions of racial slurs as expressive and descriptive. Group Processes & Intergroup Relations, 20(6), 813-830.

https://doi.org/10.1177/0261927X20904983

Sachdev, I., & Bourhis, R. Y. (1987). Status differentials and intergroup behavior. European journal of Social Psychology, 17(3), 277-293. https://doi.org/10.1002/ejsp.2420170304

Volpato, C. (2013). Negare l'altro: la deumanizzazione e le sue forme. Psicoterapia e Scienze Umane, 47(2), 311-328. https://doi.org/10.3280/PU2013-002012

Wang, C. S., Whitson, J. A., Anicich, E. M., Kray, L. J., & Galinsky, A. D. (2017). Challenge your stigma: How to reframe and revalue negative stereotypes and slurs. Current Directions in Psychological Science, 26(1), 75-80. https://doi.org/10.1177/0963721416676578

22. Objectification in the Workplace: Cognitive and Motivational Processes Undermining Workers' Humanness

Luca Andrighetto[1] *and Cristina Baldissarri*[2]
[1]University of Genoa, Italy
[2]University of Milano-Bicocca, Italy

For most of us, work is not only something we have to do to get through the month. The work we do is a means to achieve social appreciation and high status. Further, work allows us to define our human identity (Gini, 1998). In the encyclical *On Human Work,* John Paul II clarifies this issue better than any other scholar: *"through work, [woman] and man not only transform nature, adapting it to [her]/his own needs, but [she]/he also achieves fulfillment as a human being and indeed in a sense becomes more a human being".*

Thus, ideally, work is a powerful vehicle to achieve our highest representation as unique and fully human agents. At the same time, as a sort of "boomerang" effect, work may also prove to be a primary source of loss of one's own humanity, which results both in dehumanizing perceptions (and treatments) against specific occupational categories and workers' self-view as less than human. As we will discuss below, these dehumanizing processes are rooted in well-defined cognitive and motivational mechanisms and mostly target people who perform low-status jobs characterized by certain activities, such as factory workers and garbage collectors (Fiske & Dupree, 2014; Valtorta et al., 2019).

However, self-dehumanizing perceptions may also arise when performing higher status and not devalued jobs. Let's think about the academic job: in this case, the organizational climate and the daily activities

we do largely determine the perception of our job as a realization of our-selves as human beings or, inversely, as a source of erosion of our human-ness. For example, constantly feeling to be forced to do a range of repet-itive activities perceived as meaningless (i.e., the *bullshit jobs*; see Zuolo, 2022), such as quickly answering several impersonal emails or fulfilling endless and repetitive bureaucratic tasks, may enhance the researcher's sense of alienation and the view of being as a mere interchangeable tool for others' purposes. These perceptions are the core of (self-) objectifica-tion in the workplace, which is the topic of the present chapter.

Outlining the Process of Working Objectification: Theoretical and Historical insights

We conceive objectification as a specific form of dehumanization (see Baldissarri et al., 2022), that refers to the view and treatment of others – individuals or whole groups – as mere objects (Nussbaum, 1995; Vaes et al., 2014). This psychological process occurs through a gradual and often implicit erosion of others' humanness, which finally leads to considering them exclusively for their usefulness to the achievement of one's own purposes or desires (Bartky, 1990; Frederickson & Roberts, 1997; Gruen-feld et al., 2008).

Objectification has deeply shaped the past and today's work realm. Currently, the most representative image of working objectification is embodied by Amazon pickers. Their working conditions are portrayed in the BBC documentary *The truth behind the click* (2013): their daily ac-tivities are highly repetitive and mostly limited to picking orders in the warehouse, which they must execute like efficient tools. Further, their work pace is entirely imposed by a timer remote control. These work fea-tures make them feel replaceable by someone else when no longer need-ed or efficient, like interchangeable numbers (e.g., Bloodworth, 2018) and these objectifying metaphors, mostly targeting low-status workers, char-acterize all the history. Indeed, in ancient Roman times, work was con-ceived as a constriction reserved for slaves, who were defined as *animate tools* (Aristotle, trans. 1995) or *instrumenti genus vocale* (i.e., talking tools; Varro, trans. 1954). During Medieval times, manual workers were viewed as "incomplete cases of humanity" and labeled as *minus habens*. In the same vein, in the 1800s, when referring to the Afro-American slaves, an US Chief Justice talked about "an ordinary article of traffic and merchan-dise" (Dredd Scott v. Sandford, 1856). Taken together, these metaphors

effectively express the main function of this process: a powerful cognitive strategy that allows dominant groups to normalize the exploitation of workers occupying low-status positions (Volpato et al., 2017).

The objectification of workers became a highly debated issue with the rise of capitalism. In particular, Marx in his *Economic and Philosophic Manuscripts of 1844* (1844) argued that work was no longer a means through which human beings can express their humanity, but instead, an external imposition that spoils individuals of their uniquely human qualities. That is, the capitalistic system contributed to transforming workers from full human beings to mere instruments who are judged merely in terms of their productivity, rather than in terms of their humanness. However, despite its past and present pervasiveness, only recently working objectification has drawn the attention of empirical researchers. Traditionally, social psychology scholars have analyzed the process of objectification within the sexual realm. A large body of literature (for reviews, see Moradi & Huang, 2008; Guizzo, 2022 in this book; Loughnan & Pacilli, 2014) has investigated the antecedents and consequences of the male objectifying gaze toward women. Indeed, research on sexual objectification meaningfully contributed to the more recent flourishing of empirical works on objectification in the workplace, both theoretically and methodologically. One of the most important findings that inspired our line of research concerns the cognitive origins of this process. With regard to this issue, Bernard and colleagues (see Bernard et al., 2018 for a recent review) conducted a large set of studies employing cognitive or neuroscientific paradigms, by revealing that under certain conditions women are even visually elaborated like objects. That is, when women's sexualized body parts are made salient, people tend to cognitively elaborate them following an analytical path, which is typically activated for object stimuli, rather than a configurational path, which is typical for human stimuli. Thus, this evidence suggests that objectification is a phenomenon rooted in basic cognitive processes: second-order processes, such as cultural influences and beliefs about women's appearance, deeply affect people's perceptions to the point of shaping how they cognitively process objectified targets. Besides these insights, our research on working objectification has been inspired by the theoretical conceptualization of objectification provided by Nussbaum. In her sage *Objectification* (1995), she argued that five key features drive the objectified perception and treatment of workers in contemporary societies. According to her, instrumentality is the most insidious facet of this dehumanizing act: when objectified, workers are viewed

as mere instruments, and the behaviors of the "objectifier" (e.g., the superior) toward them are entirely driven by exploitation purposes. Further, workers are considered fully fungible (fungibility), that is they are seen as interchangeable with other "able-bodied workers", including robots. They are spoiled of uniquely human qualities, including the ability to plan or make decisions (denial of autonomy) or to feel emotions (denial of subjectivity). Finally, they are implicitly perceived as violable entities (violability), who for instance must be able to afford degrading working conditions that guarantee the maximum profit for the organization.

As explained later, in our line of research we conceived (and assessed) the process of working (self-)objectification by summing up the above features into two main ones (see Table 1 for a more detailed conceptualization; see also Vaes et al., 2014): instrumentality and denial of humanness, which implies a general view of the objectified target as a passive agent which is unable to feel human emotions.

Table 1. Conceptualization of working objectification from a social psychological perspective.

Working objectification: Perception of workers as objects	
Instrumentality The perception and treatment of workers as instruments	**Denial of humanness** The reduced attribution of human qualities to workers
• Useful for other purposes • Fungible with other workers • Violable and breakable • Owned by someone else	• Lacking Subjectivity (experience ·, warmth ·, human nature ·) • Lacking Autonomy (agency ·, competence ·, human uniqueness ·) • Inert

Note. The table integrates the dimensions of objectification outlined by Nussbaum (in black) with ·Dimensions of mind (Gray et al. 2007), ·Stereotype Content Model Dimensions (Fiske et al. 2007) and ·Dimensions of Humanness (Haslam, 2006).
In grey the facets of objectification that do not characterize the work domain following the Nussbaum's analysis.
From "The longstanding view of workers as objects: antecedents and consequences of working objectification", by C. Baldissarri, L. Andrighetto and C. Volpato, 2022, *European Review of Social Psychology, 33*, p. 85.

Motivational and Cognitive Predictors of Working Objectification

Like any social phenomenon, multiple factors predict working objectification. The findings obtained so far led us to think that this form of objectification is shaped by an interplay of motivational (i.e., *asymmetrical power relations* and the *salience of money*) and cognitive (i.e., *the processes activated by the performed work activities*) predictors.

First, objectifying perceptions likely arise when *asymmetrical power relations* are made salient and people occupy high-power positions. Put simply, power is the first important motivational predictor of working objectification. This assumption has been corroborated through a series of experimental studies by Gruenfeld and colleagues (2008). They found that both executives and undergraduates who occupied high-power positions were instructed to think about a high-power situation – compared to those that occupied low-power positions or were in baseline conditions – systematically objectified their work partner, by perceiving him/her as a mere instrument for the attainment of their own purposes. Further, they approached him/her only when she/he was perceived useful to these purposes. Once again, these findings make clear that objectification, and more broadly dehumanization, is a strategy that people use to justify their position and legitimize the instrumental exploitation of their subordinates.

Besides power, the *salience of money* is a second motivational predictor of objectification in the contemporary economic and working system. In fact, this salience, or the mere prospect of making money, favors the adoption of a market pricing mode, a motivational mindset based on a rational calculation of the self-gain deriving from interpersonal relations (Fiske, 1991). In turn, this mindset is a breeding ground for objectifying tendencies when interacting with others, especially when they are viewed as useful for the achievement of gain. Accordingly, Teng and colleagues (2016, Experiment 1) for example demonstrated that undergraduates who were asked to depict coins or banknotes (condition of the salience of money) reported higher levels of market pricing mindset and, in turn, reported higher levels of objectifying tendencies during interpersonal relations than undergraduates who depicted furniture (control condition). Further, Wang and Krumhuber (2016; Study 2) reported that participants with an activated affective state of being rich displayed increased objectifying tendencies toward a work partner, who was treated in an instrumental way and judged based on goal contingent than personal traits.

Asymmetrical power relations and the prospect of making money are undoubtedly two key antecedents of objectification in the workplace. However, we argued that objectifying perceptions against workers could also emerge in the absence of these two motivational forces and be embodied in the work activity itself. We conducted a set of studies (Andrighetto et al., 2017; Baldissarri et al., 2017) to demonstrate that the work

activities that an individual performs in certain workplaces represent per se an important cognitive source of laypeople's objectified perceptions. In doing so, we integrated experimental paradigms from sexual objectification literature (e.g., Heflick & Goldenberg, 2009) with classical sociological analyses (Blauner, 1964), that pointed out that three features of the factory job are especially linked to workers' alienation and to the view of workers as things: the repetitiveness of movements, the fragmentation of activities, and the other-directed rhythm of pace. In our studies (see Andrighetto et al., 2017), we made salient undergraduates these features by for example exposing them to pre-tested video clips that portrayed a factory worker while performing his daily activities. Further, during the view, we asked them to focus on his work or on him as a person, depending on the experimental condition. Overall, our findings revealed that participants who focused on the work perceived the target as more instrument-like and denied him more human mental states than those who were asked to focus on his personhood. Instead, these differences did not emerge for participants who were exposed to a clip depicting another manual laborer, i.e., the artisan, as the features of this work do not convey an objectified view of the worker (see Figure 2). We expanded these results with a subsequent study (Baldissarri et al., 2017) in which we employed a similar procedure but assessed the objectifying perceptions through an Implicit Association Test (Greenwald et al., 1998).

Attributions of mental states

Figure 1. Attributions of mental states as a function of target (factory worker vs. artisan) and focus (work vs. personhood) manipulation. Andrighetto et al. 2017 (Study 3)
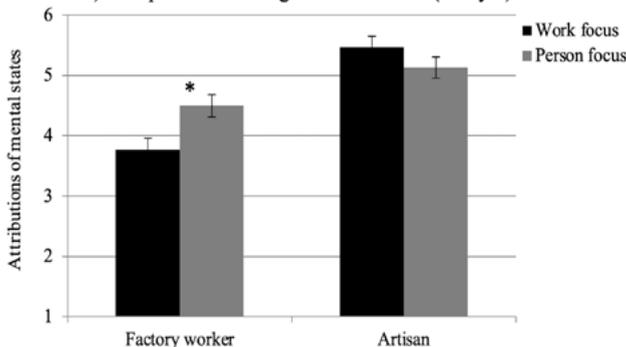
Note. From "(Still) Modern Times: Objectification at work", by L. Andrighetto, C. Baldissarri, and C. Volpato, 2017, *European Journal of Social Psychology, 47,* p. 31.

Consistent with the previous studies, findings revealed that focusing on the repetitive, fragmented and other-directed activities performed by the factory worker (vs. the artisan) led participants to implicitly associate him with object- rather than person-related stimuli. This latter finding further strengthens the idea that working objectification may also be a consequence of a cognitive process, in which workers performing certain activities are automatically assimilated into mere objects. More specifically, these results led us to think that the salience of these objectifying work features triggers a process of inductive inference per se, leading people to perceive the human targets who perform them as nonhuman entities. This assumption is also supported well by intriguing research by Maass and colleagues (2001), who showed that people base their social inferences mainly and more spontaneously on a behavior-to-trait induction process, attributing traits (or not attributing them as in the case of workers' objectification) based on behavioral information that they collect. Further, this inference could be conceived as symmetrical to the cognitive process that triggers the anthropomorphism of nonhuman agents (Epley et al., 2007). As social robots with anthropomorphic movements are assimilated to humans and attributed uniquely human abilities, human beings who perform repetitive, fragmented and other-directed actions may be perceived as mindless and passive instruments and, thus, stripped of their uniquely human attributes.

On the intra- and inter-personal consequences of working objectification

Our investigation on working objectification examined in parallel the antecedents and the possible consequences of this process for the objectified target. Experiencing objectification in the workplace may have a wide range of detrimental outcomes, affecting both personal and societal levels.

Similar to sexual objectification (see Fredrickson & Roberts, 1997; Guizzo, 2022 in this book), the internalization of the objectifying gaze is the primary intrapersonal consequence of this dehumanizing process. Being instrumentally *objectified by one's own superior* (Baldissarri et al., 2014) or being treated by the organization as interchangeable instruments for making profits (i.e., *organizational dehumanization*; Caesens et al., 2017) could lead workers to assimilate these perspectives by eliciting

self-objectification, both in terms of the self-view as more instrument- than human-like and as unable to feel human mental states. However, *performing objectifying job activities* is per se a further source of workers' self-objectification. We first demonstrated this latter assumption experimentally, through a set of laboratory studies. In these experiments, we created ad hoc tasks in which we manipulated the key objectifying job features (e.g., Baldissarri et al., 2017). In particular, we validated the "ACME shop" paradigm (see Baldissarri et al., 2020), a simulation of a computer job activity in which participants were asked to role-play a computer shop online seller. In the objectifying condition, the task was created to be highly repetitive, fragmented (i.e., they had just to select the requested product) and other-directed (i.e., a time at the top of the screen set the pace of their work). Instead, in the non-objectifying condition, participants performed different tasks (e.g., compiling the package, managing appointments with the customers) without any indications about the pace of their work. Results showed that participants assigned to the objectifying task self-attributed fewer human mental states to themselves and self-perceived, both explicitly and implicitly, as more instrument-like (vs. human-like) than participants who simulated the job activity in the non-objectifying condition or participants in a baseline condition (i.e., a ludic activity). We then replicated these findings outside the laboratory, through cross-sectional research that involved a large sample of workers (e.g., metalworkers, grocery clerks) employed by different Italian factories or organizations (see Baldissarri et al., 2022). Overall, these studies revealed that the greater the feeling of performing a repetitive, fragmented and other-directed job activity, the higher the workers' self-objectification, both in terms of instrumentality and denial of humanness. These results hold also when controlling for other possible antecedents, such as the perception of being objectified by their superiors. A very recent study (Baldissarri et al., 2022) provided evidence for a further antecedent of self-objectification. Through four studies also involving permanent or temporary workers, we found that the *perceived job insecurity* – one of the main stressors for nowadays workers – is positively associated with people's self-view as objects, who thus perceive being at the mercy of the event and external forces.

The increased self-objectification due to objectifying treatments, specific job features, or a sense of precariousness is then a fundamental trigger of a series of further consequences for the individual (see Figure 3). First, self-objectified workers are more likely to display increased burn-

out (Baldissarri et al., 2014), undermined well-being (Baldissarri et al., 2022), or less engagement toward the job task (Baldissarri et al., 2021). However, this state is also associated with social psychological outcomes that are of great relevance to the nowadays societies. More specifically, we provided consistent evidence that this increased self-perception as objects, especially due to performing objectifying activities (Baldissarri et al., 2017), leads people to believe having less free will (Baldissarri et al., 2017, 2019) and to a greater propensity to conform to others (Andrighetto et al., 2018; Baldissarri et al., 2020). That is, it is plausible to imagine that this internalized state of being a passive agent unconsciously implies a decreased sense of autonomy and conscious choice, which is then translated into a decreased sense of having free will and more conformity. We argue that especially these latter consequences could gain importance at a more societal level and somewhat contribute to explaining the existing social inequalities (see Bruckmüller, 2022, in this book). We refer particularly to free will: its undermining inhibits people from hoping that they can improve their situation and, thus, more likely to accept it in a passive way (Baumeister & Monroe, 2014). As a consequence, this belief may impact the active societal role of low-status workers who are more likely exposed to objectifying treatments or activities, deterring their tendency to engage in collective actions to change the existing status quo. Very preliminary evidence seems to support this idea (Baldissarri et al., 2022): in a study that considered a sample of Italian factory workers, we indeed found that their feelings of self-objectification were correlated with reduced beliefs in having free will and, in turn, with increased tendencies to justify the system and decreased activist tendencies.

Figure 2. Main antecedents and consequences of workers' self-objectification

Antecedents

- Objectification by one's own superior
- Organizational dehumanization
- Performing objectifying activities
- Perception of job insecurity

▼

Self-objectification

- Self-view as instrument- (vs- human-like)
- Self-denial of human mental states

▼

Intra-and inter-personal consequences

- Well-being
- Burnout
- Engagement with the task

- Belief in free will
- Conformity

▼

Social consequences

- Collective action
- System justification

Is it possible to reduce working objectification?

The literature reviewed above clearly indicates that objectification in the workplace is a persistent phenomenon over time, with potentially harsh consequences both at an individual and societal level. Thus, there is an impelling need to identify strategies that could mitigate its pervasiveness and effects. Nevertheless, it is important to point out that in the last few decades there has been a significant increase in organizations that have put great efforts into improving the working conditions of their employees, by first protecting and preserving their human dimension. Notably, many of these initiatives were inspired by enlightened approaches from organizational psychology, based for example on the Human Relations Management Theory (Mayo, 1945). However, much more work needs to be done, both at a theoretical and more applied level. In particular, the individual and organizational factors preventing workers from objectification and its effects still miss a systematic investigation. However, some research groups are providing first precious insights in this direction. For example, Auzoult and Personnaz (2016) reported that individual self-consciousness is an important self-regulation process that might prevent individuals from internalizing the external objectifying gaze. Further, Caesens and colleagues (2017) reported that the perception of being supported by the organization is a key variable in mitigating the workers' perceptions of being dehumanized by their organization. Even more interestingly, recent research by Teresi and colleagues (2022) documented that employees' self-objectification is significantly reduced when they feel that their workplace is characterized by an ethical climate of interdependence and, in turn, self-identify with the organization.

We hope that these and upcoming findings are then translated into concrete guidelines and policies that could prevent working objectification, at the same time contributing to the creation of work environments that recognize the human needs of each worker.

References

Andrighetto, L., Baldissarri, C., Gabbiadini, A., Sacino, A, Valtorta, R.R., & Volpato, C. (2018). Objectified conformity: Working self-objectification increases conformist behaviour. *Social Influence*, *13*, 78-90. doi: 10.1080/15534510.2018.1439769
Andrighetto, L., Baldissarri, C., & Volpato, C. (2017). (Still) modern times:

Objectification at work. *European Journal of Social Psychology, 47,* 25–35. doi: 10.1002/ejsp.2190

Aristotle (1995). *Aristotle: Selections.* Trans. T. Irwin, & G. Fine. Indianapolis: Hackett

Auzoult, L., & Personnaz, B. (2016). The role of organizational culture and self-consciousness in self-objectification in the workplace. *TPM, Testing, Psychometrics, Methodology in Applied Psychology, 23*, 1–14. doi:10.4473/TPM23.3.1

Baldissarri, C., & Andrighetto, L. (2021). Being treated as an instrument: Consequences of instrumental treatment and self-objectification on task engagement and performance. *Human Performance, 2,* 85-106. doi: 10.1080/08959285.2021.1878182

Baldissarri, C., Andrighetto, L., Di Bernardo, G.A., & Annoni, A. (2020). Workers' Self-objectification and Tendencies to Conform to Others. *Journal of Community and Applied Social Psychology, 30,* 547-560. doi: 0.1002/casp.2461

Baldissarri, C., Andrighetto, L., Gabbiadini, A., & Volpato, C. (2017). Work and freedom: Working self-objectification and belief in personal free will. *British Journal of Social Psychology, 56,* 250-269. doi:10.1111/bjso.12172

Baldissarri, C., Andrighetto, L., Orsenigo, F., & Volpato, C. (2022). Workers' self-objectification and reduced activism against inequalities: The role of beliefs in personal free will and system justification. *Psicologia Sociale.* Manuscript accepted for publication.

Baldissarri, C., Andrighetto, L., & Volpato, C. (2014). When work does not ennoble man: Psychological consequences of working objectification. *TPM, Testing, Psychometrics Methodology in Applied Psychology, 21(3),* 327-339. doi: 10.4473/TPM21.3.7

Baldissarri, C., Andrighetto, L., & Volpato, C. (2019). Feeling like an object: A field study on working self-objectification and the belief in personal free will. *TPM, Testing, Psychometrics Methodology in Applied Psychology, 26,* 185-197. doi: 10.4473/TPM26.2.

Baldissarri, C., Andrighetto, L., & Volpato, C. (2022). The longstanding view of workers as objects: antecedents and consequences of working objectification, *European Review of Social Psychology, 33*, 81-130, doi: 10.1080/10463283.2021.1956778

Baldissarri, C., Gabbiadini, A., Andrighetto, L., & Volpato, C. (2020). The ACME Shop: A paradigm to investigate working (self-) objectification. *The Journal of Social Psychology*, 1-17.

Baldissarri, C., Valtorta, R.R., Andrighetto, L., & Volpato, C. (2017).

Workers as objects: The nature of working objectification and the role of perceived alienation. *TPM, Testing, in Applied Psychology, 21(3),* 327-339. doi: 10.4473/TPM21.3.7

Baldissarri, C., Teresi, M., Pagliaro, S., & Andrighetto, L. (2022). Humanness in times of uncertainty: The impact of perceived job insecurity on self-objectification and well-being. *European Journal of Social Psychology.* Manuscript accepted for publication.

Bartky, S. (1990). *Femininity and domination: Studies in the phenomenology of oppression.* New York: Routledge Press.

Baumeister, R. F., & Monroe, A. E. (2014). Recent research on free will: conceptualizations, beliefs, and processes. *Advances In Experimental Social Psychology*, 50, 1–52. doi:10.1016/B978-0-12-800284-1.00001-1

BBC (2013, November 23). *The truth behind the click.* Retrieved from www.bbc.co.uk/programmes/b03k5zp.

Bernard, P., Gervais, S. J., & Klein, O. (2018). Objectifyng objectification: When and why people are cognitively reduced to their parts akin to objects. *European Review of Social Psychology, 29, 82-121.* doi: 10.1080/10463283.2018.1471949.

Blauner, B. (1964). *Alienation and Freedom: The Factory Worker and His Industry.* Chicago, IL: University of Chicago Press.

Bloodworth, J. (2018). *Hired: Six Months Undercover in Low Wage Britain.* London: Atlantic Books.

Bruckmüller, S. (2022). Inequality as privilege versus disadvantage: How and why that matters.

Caesens, G., Stinglhamber, F., Demoulin, S., & De Wilde, M. (2017). Perceived organizational support and employees' wellbeing: The mediating role of organizational dehumanization. *European Journal of Work and Organizational Psychology, 26,* 527–540. doi: 10.1080/1359432X.2017.1319817

Epley, N., Waytz, A., & Cacioppo, J. T. (2007). On seeing human: a three-factor theory of anthropomorphism. *Psychological Review, 114*, 864-886.

Fiske, A. P. (1991). *Structures of social life: The four elementary forms of human relations: Communal sharing, authority ranking, equality matching, market pricing.* Free Press.

Fiske, S. T., Cuddy, A. J. C., & Glick, P. (2007). Universal dimensions of social cognition: Warmth and competence. *Trends in Cognitive Sciences, 11*, 77–83. Doi: 10.1016/j.tics.2006.11.005

Fiske, S. T., & Dupree, C. (2014). Gaining trust as well as respect in communicating to motivated audiences about science topics.

Proceedings of the National Academy of Sciences, 111(supplement_4), 13593-13597.

Fredrickson, B. L., & Roberts, T. A. (1997). Objectification theory: Toward understanding women's lived experiences and mental health risks. *Psychology of Women Quarterly, 21,* 173-206. doi:10.1111/j.1471-6402.1997.tb00108.x

Gini, A. (1998). Work, Identity and Self: How We Are Formed by The Work We Do. *Journal of Business Ethics, 17,* 707–714. https://doi.org/10.1023/A:1017967009252

Gray, H. M., Gray, K., & Wegner, D. M. (2007). Dimensions of mind perception. *Science, 315,* 619. doi: 10.1126/science.1134475.

Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. (1998). Measuring individual differences in implicit cognition: the implicit association test. *Journal of Personality and Social Psychology, 74,* 1464.

Gruenfeld, D. H., Inesi, M. E., Magee, J. C., & Galinsky, A. D. (2008). Power and the objectification of social targets. *Journal of Personality and Social Psychology, 95,* 111–127. doi: 10.1037/0022-3514.95.1.111

Guizzo, F. (2022). Sexual objectification: consequences and counteractions

Haslam, N. (2006). Dehumanization: An integrative review. *Personality and Social Psychology Review, 10,* 252-264. doi: 10.1207/s15327957pspr1003_4

Heflick, N. A., & Goldenberg, J. L. (2009). Objectifying Sarah Palin: Evidence that objectification causes women to be perceived as less competent and less fully human. *Journal of Experimental Social Psychology, 45,* 598–601. https://doi.org/10.1016/j.jesp.2009.02.008

Loughnan, S., Haslam, N., Sutton, R. M., & Spencer, B. (2014). Dehumanization and social class: Animality in the stereotypes of "White Trash," "Chavs," and "Bogans". *Social Psychology, 45,* 54-61. Doi: 10.1027/1864-9335/a000159

Loughnan, S., & Pacilli, M. G. (2014). Seeing (and treating) others as sexual objects: toward a more complete mapping of sexual objectification. *TPM: Testing, Psychometrics, Methodology in Applied Psychology, 21,* 309-325.

Maass, A., Colombo, A., Colombo, A., & Sherman, S. J. (2001). Inferring traits from behaviors versus behaviors from traits: The induction–deduction asymmetry. *Journal of Personality and Social Psychology, 81,* 391-404. Doi: 10.1037/0022-3514.81.3.391

Marx, K. (1844). *Economic and Philosophic Manuscripts of 1844.* Trans. M. Milligan. Mineola, NY: Dover Publications, Inc., 2012.

Mayo, E. (1945). *The social problem of an industrial civilization.* Boston:

Harvard Business School.

Moradi, B., & Huang, Y. (2008). Objectification theory and psychology of women: A decade of advances and future directions. *Psychology of Women Quarterly, 32,* 377-398. doi:10.1111/j.1471-6402.2008.00452.x

Nussbaum, M. (1995). Objectification. *Philosophy and Public Affairs, 24,* 249-291. doi:10.1111/j.1088-4963.1995.tb00032.x

Teng, F., Chen, Z., Poon, K. T., Zhang, D., & Jiang, Y. (2016). Money and relationships: When and why thinking about money leads people to approach others. *Organizational Behavior and Human Decision Processes, 137,* 58-70. doi: 10.1016/j.obhdp.2016.08.002

Teresi, M., Ballone, C., Barattucci, M., Baldissarri, C., Andrighetto, L., & Pagliaro, S. (2022). Examining workers' self-objectification through the lens of social identity: The role of ethical climate and organizational identification. *Psicologia Sociale, 17,* 93-102.

Vaes, J., Loughnan, S., & Puvia, E. (2014). The inhuman body: When sexual objectification become dehumanizing. In P. Bain, J. Vaes, & J.-Ph. Leyens (Eds.), *Humanness and dehumanization* (pp. 186-204). Hove, UK: Psychology Press.

Valtorta, R. R., Baldissarri, C., Andrighetto, L., & Volpato, C. (2019). Dirty jobs and dehumanization of workers. *British Journal of Social Psychology, 58,* 955-970. doi: 10.1111/bjso.12315

Volpato, C., Andrighetto, L., & Baldissarri, C. (2017). Perceptions of low-status workers and the maintenance of the status quo. *Journal of Social Issues, 73,* 192-210. doi: 10.1111/josi.12211

Zuolo, F. (2022). *"Bullshit jobs" in università: un'autodenuncia [Bullshit jobs in the academic world: A self-declaration].* Retrieved from https://www.valigiablu.it/burocrazia-universita/

23. Sexual objectification: an introduction

Francesca Guizzo
University of Padova, Italy

"What is objectification? I heard this word a lot of times, for example, "women shouldn't be objectified" but never really understood what it means?" writes a user on Quora. Many other users echo this question, while others give their take on what it means to be objectified. This example shows that the terms objectification and sexual objectification have entered common usage, but many people may not be aware of their meaning, not to mention their psychological implications. The purpose of this chapter is to reduce this lacuna by providing a non-comprehensive overview of the psychological research on sexual objectification. I will start by defining sexual objectification and explaining its philosophical and psychological roots. Then, I will introduce objectification theory (Fredrickson & Roberts, 1997), which describes the detrimental effects of sexual objectification on women's self-perception (self-objectification) and well-being. The chapter also covers research on how women are viewed (dehumanized) and treated by others when sexually objectified, including one of the most dangerous and extreme consequences, namely sexual harassment and violence. Finally, I will briefly discuss possible ways to prevent and address sexual objectification.

What is sexual objectification?

The concept of objectification has a long history in Philosophy. Immanuel Kant was the first to describe it in 1785 (1963) as the process of reducing a person to an object, treating them as a means to an end, and stripping them of their dignity. This concept has since been revisited by

many philosophers and thinkers, including Marx (1844, 1964). Notably, Martha Nussbaum (1995) has expanded on the idea of objectification by suggesting that there are seven ways to objectify a person (p. 257). Specifically, a person is objectified whenever they are subjected to one or more of the following behaviors:

- *Instrumentality*: a person is treated as a tool for one's end.
- *Denial of autonomy*: a person is regarded as "lacking self-determination".
- *Inertness*: a person is treated as if they lack the ability to control their actions and their consequences (i.e., agency).
- *Fungibility*: a person is treated as if they are interchangeable with other persons/objects.
- *Violability*: a person can be broken up or smashed as they have no "boundary integrity".
- *Ownership*: a person is treated as a commodity that can be bought or sold.
- *Denial of subjectivity*: A person's feelings and experiences are denied.

Although objectification may be enacted toward any individuals and in different contexts (see for example the objectification of workers; Andrighetto & Baldissari in this book; Baldissarri et al., 2022), feminist thinkers across different disciplines (e.g., Bartky, 1990; de Beauvoir, 1952; Fredrickson & Roberts, 1997; McKinnon, 1993; Nussbaum, 1995) observed that women are more often the targets of objectifying treatments, which assume a specific form called sexual objectification. In fact, women are, more frequently than men, judged only on the basis of their physical appearance (Bartky, 1990), which is the essence of sexual objectification. Specifically, Sandra Bartky (1990) contends that sexual objectification is a *fragmentation* process whereby women are identified with their bodies or sexual body parts that are separated from their personhood becoming mere instruments for the pleasure of others (Fredrickson & Roberts, 1997).

Taking this definition into account, Langton (2008) extended Nussbaum's objectifying conducts (1995) by adding other three features that occur whenever sexual objectification is perpetrated, namely *reduction to body* (individuals are identified with their body or body parts), *reduction to appearance* (individuals are evaluated primarily in terms of how they appear) and *silencing* (individuals lack the ability to speak). Altogether, sexual objectification can be defined as a form of reduction to the body

that occurs whenever people (more often women) are fragmented into a collection of sexual body parts or functions, considered as silent decorations, and evaluated solely based on their appearance, while their personality and other qualities are devalued.

More recently, the term sexualization has been introduced. This is an umbrella concept that comprises any instances/experiences where sexuality is imposed upon a person, or when being attractive is narrowly reconducted to being sexy, the individual is valued only on their sexual appeal, or the person is sexually objectified as defined above (APA, 2007; see Ward, 2016, Roberts et al., 2018 for a discussion). It is important to note that not all expressions of sexuality or sexual attraction are inherently harmful. For example, Nussbaum (1995) argued that sexual objectification may be harmless in the context of mutual sexual interest within romantic relationships. However, sexualization and sexual objectification become harmful when are used to diminish the person and when linked with dehumanization (see Pecini et al., 2023; Vaes et al., 2013 for a discussion), a concept that will be reviewed more thoroughly below.

Sexual objectification, and sexualization more broadly, permeates women's life, especially in Western societies (APA, 2007; Fredrickson & Roberts, 1997). Examples of everyday manifestations of sexual objectification include being ogled, observing other women being sexualized (e.g., in the media), or even suffering more extreme experiences of sexual harassment (e.g., Holland et al., 2017; Swim et al., 2001). Despite the prevalence of objectification in women's lives, it was not until Fredrickson and Roberts published objectification theory in 1997 that social psychologists began to systematically investigate its psychological effects on women. Objectification theory is a social psychological framework that explains how women's bodies and appearance are objectified and sexualized in society, leading to negative consequences for women's mental and physical health, as well as their ability to fully participate in society. We will now review the objectification theory tenets.

Objectification theory

Building on feminist philosophers' arguments, Fredrickson and Roberts (1997) first analyzed how sexual objectification is perpetrated. They proposed that perhaps the most powerful way in which women are sexually objectified is the sexually objectifying gaze (i.e., visual inspection of the body) because it subtly conveys to women the message that they

are being evaluated based on their body appearance. They further argued that interpersonal encounters and visual media are the two main contexts in which sexual objectification is played out. Women of all ages, indeed, report experiencing sexually objectifying interactions almost daily, including having their bodies scrutinized, receiving catcalls, and extreme forms of unwanted sexual advances (e.g., Holland et al., 2017; Kozee et al., 2007; Swim et al., 2001). It is also important to note that women report experiencing sexually objectifying interactions to a greater extent than men (see Swim et al., 2001).

In addition, women are exposed to sexual objectification every day via mass media (Holland et al., 2017). Both traditional and social media often emphasize women's bodies and appearance, subtly aligning the viewers with a sexually objectifying gaze. As a matter of fact, very often we come across images of scantily dressed women, framed in sexy and provocative poses that seem to have the only purpose to increase audience and consumers (Ward, 2016). Men are not excluded from such treatment (e.g., Carrotte et al., 2017); however, women are more likely than men to be depicted in sexualized ways, for instance in advertisements, magazines, films, TV programs, music videos as well as in social media (e.g., Archer et al., 1983; Aubrey & Frisby, 2011; Carrotte et al., 2017; Hatton & Trautner, 2011; Smith et al., 2019; Vandenbosch et al., 2013; see Galdi & Guizzo, 2021 and Ward, 2016 for reviews). Altogether this evidence demonstrates that women and girls are the most frequent targets of sexual objectification during both interpersonal encounters with familiar people or strangers, and in visual media.

From Sexual Objectification to Self-Objectification

Living in a context in which the female body is constantly scrutinized, and the value of physical appearance over other qualities is reinforced, has dreadful repercussions on girls and women. Women start to learn from an early age that being sexually attractive is a crucial aspect of being a woman (Fredrickson & Roberts, 1997). As a consequence, women learn to self-objectify, that is to value themselves as objects that exist primarily to be looked at and evaluated by others. In other words, women and girls adopt the same observer's gaze on themselves, therefore assuming a third-person perspective that leads them to value themselves mostly in terms of how their body appears to others (Fredrickson & Roberts, 1997).

Fredrickson and Roberts (1997) further argued that the objectifying gaze interiorized by women is especially the male gaze. Western cultures

are based on a heteronormative and patriarchal framework (see Bareket & Shnabel, 2020 for a discussion), which often "measures women's value in relation to their fulfillment of the role of sex object for men" (Calogero, 2013a, p. 99) and where sexual objectification may be used by men to reinforce their dominance over women (Bareket & Shnabel, 2020). Although evidence shows that women objectify other women (e.g., Strelan & Hargreaves, 2005), women doing so are likely taking the male gaze perspective because this is what they are socialized to do (Calogero, 2013a). To summarize, objectification theory proposes that an important repercussion of being repeatedly valued on the sole basis of physical appearance standards, shaped by cultural demands to be sexually attractive to men, is that girls and women learn, over time, to internalize such observer's perspective on the self, a process that scholars have named self-objectification (Fredrickson & Roberts, 1997).

Self-objectification has been conceptualized both as a trait disposition to chronically view oneself as a body, or a situational state that could be triggered by a sexually objectifying situation, such as having the body scrutinized or being catcalled. In experimental settings, state self-objectification has been triggered, for example, by asking participants to wear a swimsuit (vs. jumper) in front of a mirror, having them watching other women being sexually objectified in the media, or have the experimenter film or take pictures of them from the neck down (see Kahalon et al., 2018 for a review). Moreover, self-objectification has also been operationalized in different ways. Typically, it has been measured as the difference between the perceived importance of body appearance versus body competence (Fredrickson et al., 1998), or as the degree of body surveillance, which refers to the extent to which an individual monitors their own body and appearance (McKinley & Hyde, 1996; see Kahalon et al., 2018 for a discussion).

Whether in its trait form or in a situational state triggered by contextual objectifying experiences, self-objectification is thought to predict several adverse outcomes for women's psychological and cognitive well-being (Fredrickson & Roberts, 1997). I will now discuss the main consequences.

Consequences of Self-Objectification

Objectification theory conceptualized self-objectification as the major conjunctive mechanism between women's sexual objectification experiences at the cultural level and their psychological well-being (Fredrick-

son & Roberts, 1997). Although men and boys can self-objectify (Karsay et al., 2018) and report increased body concerns as a consequence of sexualized media exposure (e.g., Aubrey, 2006; Vandenbosch, & Eggermont, 2013; but see Daniel & Bridges, 2010 for mixed results), women and girls have been shown to suffer a disproportionate number of negative consequences. Extensive literature supports this claim and published reviews are available (e.g., Kahalon et al. 2018; Moradi & Huang, 2008; Tiggemann, 2011; Roberts et al., 2018). Therefore, I will limit myself to briefly reviewing the major links that have been tested.

Originally, Fredrickson and Roberts (1997) proposed that self-objectification, manifested as body surveillance, would directly promote a range of negative psychological consequences for women. Specifically, they proposed that it would increase body shame (resulting from one's appearance failing to meet the internalized cultural standard), as well as both appearance anxiety (i.e., the anticipation of the fear of having the body evaluated) and safety anxiety (resulting from the potential aggressive intent of the perpetrator); it would also decrease women's peak motivational states (or flow experience, i.e., rare moments of complete immersion on a task, associated with joy and pleasure) and awareness of internal bodily states (i.e., ability to detect one's internal physiological sensation, e.g., hunger, fatigue, emotions, physical sensations). They also posited that this chain of relations would ultimately impact women's mental health by increasing their risk of developing eating disorders, depressive symptoms, and sexual dysfunctions (Fredrickson & Roberts, 1997). Figure 1 summarizes the key relations proposed by the original objectification theory framework.

Throughout more than twenty decades of research, the proposed framework has been tested both in correlational and experimental studies. Specifically, experimental and correlational evidence supports the link between self-objectification and increased body shame and appearance anxiety (e.g., Fredrickson et al., 1998; Quinn et al., 2006). In turn, increased body shame and appearance anxiety has been linked to depressive symptoms, restrained eating and eating disorders, as well as sexual dissatisfaction and dysfunctions (e.g., Tiggemann & Williams, 2012; Vencill et al., 2015; see Jones & Griffiths, 2015 for a review on depression and Tiggemann, 2013 for a review on eating disorders; see Calogero et al., 2021 for some inconsistent results).

The effects of self-objectification on flow experience, internal bodily awareness, and safety anxiety have received far less attention (Kahalon

et al., 2018). Data supports the association between self-objectification and reduced internal body awareness, flow experience as well as increased safety anxiety (e.g., Ainley & Tsakiris, 2013; Calogero et al., 2021; Greenleaf, 2005). However, their hypothesized ripple effects on mental health outcomes have not been completely supported or the literature is mixed (e.g., Daubenmier, 2005; Greenleaf, 2005; Tiggemann & Kuring, 2004; Tiggemann & Williams, 2012; see Calogero et al., 2021 for a recent discussion). Taken together the literature largely supports the original objectification theory framework (Figure 1), although refinements may be needed as not all the pathways among psychological experiences and mental health outcomes are supported (e.g., the link between interoceptive awareness and health outcomes; see Calogero et al., 2021).

It is important to mention that the objectification theory model has been tested in different samples, including sexual minorities women (e.g., Moradi & Tebbe, 2022), ethnic minorities women (e.g., Davies et al., 2021; Kilpela et al., 2019), and female adolescents (see Daniels et al., 2020 for a review). Results suggest that living in a society that sexually objectifies women can lead, with minor differences, both women and girls across different ages, sexual orientations, and ethnicities to experience the negative consequences of self-objectification.

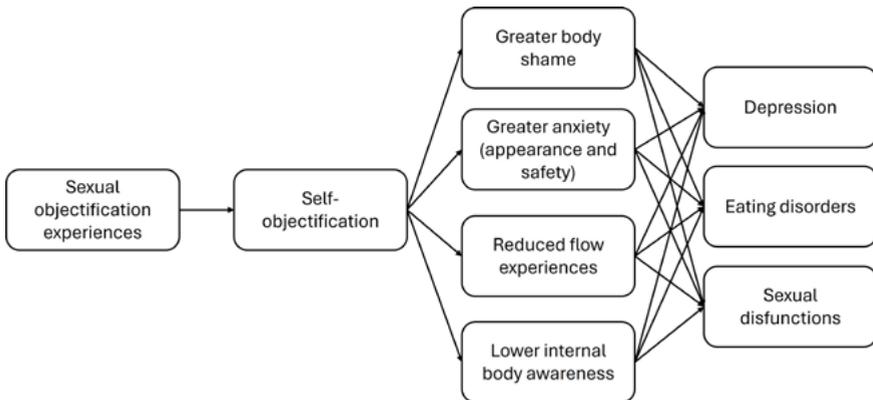Figure 1. Objectification theory framework (see also, Moradi & Huang, 2008)

As highlighted above, the objectification theoretical framework posits that sexually objectifying experiences are the central precursors to self-objectification and subsequent psychological effects (Figure 1; Fredrickson & Roberts, 1997). While a considerable amount of research on this topic is correlational, evidence exists linking exposure to sexualized

media to increased self-objectification, body shame, appearance anxiety, and body concerns among women and girls (e.g., Aubrey, 2006; Grabe et al., 2008; Slater & Tiggemann, 2016; see Karsay, 2020 and Ward, 2016 for reviews; Karsay et al., 2018 for a meta-analysis). Additionally, daily diary studies and correlational studies demonstrate that women report being frequently targeted by sexual objectification in their interactions (e.g., sexually objectifying gazes and even sexually harassing instances) and this is associated with greater self-objectification and body shame (e.g., Calogero et al., 2021; Holland et al., 2017; Koval et al., 2019; Kozee et al., 2007). Anticipation of receiving a male objectifying gaze during an interpersonal interaction also increases women's body shame and appearance anxiety (Calogero, 2004).

Besides the original psychological chain predicted by objectification theory reported above, self-objectification and sexually objectifying experiences have been found to be related to other several negative outcomes. For example, sexual objectification decreases women's cognitive performance (see Winn & Cornelius, 2020 for a review), leads women to talk less (Saguy et al., 2010), and may disrupt their willingness to participate in social activism (Calogero, 2013; but see De Wilde et al., 2020 for null effects). Moreover, sexual objectification is linked to increased interest in cosmetic surgery (e.g., Guizzo et al., 2021; Vaughan-Turnbull & Lewis, 2015) and increased menstrual shame and risky sexual behaviours (e.g., Hirscham et al., 2006; Impett et al., 2006). Furthermore, self-objectification is directly associated to lower self-efficacy and self-esteem (e.g., Gapinski et al., 2003; Mercurio & Landry, 2008), and indirectly linked to lower salary negotiation via lower self-attribution of competence (Guizzo et al., 2024). All in all, sexual objectification changes the way in which women perceive themselves and is linked to many harmful consequences for their psychological and health well-being.

Does sexual objectification impact the way in which women are perceived?

Social psychologists have only recently started to study how the observers perceive women when they sexually objectify them. They asked themselves whether sexual objectification can be seen just as a metaphor or might actually change the way in which the targets are cognitively perceived. In line with the definition, sexualized women (i.e., scantily dressed, sensual poses) are visually processed in piecemeal ways, re-

sembling the recognition of objects (Bernard et al., 2012; Gervais et al., 2012; see Bernard et al., 2018 for a review). Likewise, when individuals are instructed to focus on a woman's appearance (compared to a man's appearance), they attribute to her less competence, sociability and morality, which are fundamental characteristics of social perception (Heflick & Goldenberg, 2009; Heflick, et al., 2011). Moreover, sexualized (vs. non-sexualized) women are attributed less mind and moral status (e.g., Loughnan et al., 2010), less agency (e.g., Cikara et al., 2010; Gray et al., 2011), are more quickly associated with animal than human attributes (Vaes et al., 2011), and are perceived similarly to objects than human beings (Vaes et al., 2019; see also Vaes et al., 2020 for other results). In other words, when sexualized, women are dehumanized, namely, they are perceived as being less than humans (see Bernard et al., 2018; Heflick & Goldenberg, 2014; Pecini et al., 2023 for reviews).

Taken together, research supports the notion that sexual objectification is not only a philosophical construct, but it is also a psychological process that affects how women are cognitively and morally perceived. It should be noted that in most of the studies (e.g., Bernard et al., 2012; Gervais, et al., 2012; Heflick et al., 2011; Vaes et al., 2020), participants' gender was not a significant factor, suggesting that both men and women put in place female sexual objectification, although for different reasons. Women have been found to dehumanize other sexualized women the higher their motivation to look attractive to men and their internalization of the sociocultural beauty standards (Puvia & Vaes, 2013). This is in line with the idea that women objectifying other women may have interiorized the male gaze perspective (Calogero, 2013a). On the other hand, men may dehumanize sexualized female targets to a greater extent when they are highly attracted to those targets (Vaes et al., 2011).

Sexual Harassment as a Consequence of Sexual Objectification

Sexual harassment is a widespread phenomenon, with approximately 50% of European women reporting at least one occurrence of sexual harassment perpetrated by a man in their life (European Union Agency for Fundamental Rights- FRA, 2014; see also Stop Street Harassment, 2018 for US data). Understanding the causes of sexual harassment is crucial to enable the creation of interventions aimed at reducing this dreadful phenomenon. For example, research suggests that unwanted sexual attention (e.g., unwanted touching or comments) and sexual coercion (e.g., bribes

or threats seeking sexual favours often exerted in working settings) typically stem from the desire for sex and the need for power and dominance (Maass et al., 2013; Pryor et al., 1993). On the other hand, gender harassment, which aims to offend and demean women (e.g., sexist jokes), may arise as a response to threats to men's gender identity or as a way to enhance their masculinity, as shown by pivotal works by Anne Maass and colleagues (Dall'Ara & Maass, 2000; Maass et al., 2003).

The sexually objectifying society we live in is also considered an important risk factor (Gervais & Eagan, 2017; Maass et al., 2013). As discussed above, sexual objectification contributes to a culture in which women can be treated as objects to be exploited for sexual pleasure. This create an environment in which sexual harassment is justified. In fact, men who are more likely to dehumanize and see women as objects or animals are also more likely to engage in unwanted sexual behaviours and aggression (e.g., Bevens & Loughnan, 2019; Gervais & Eagan, 2017; Rudman & Mescher, 2012; see also Vasquez et al., 2018 for experimental data).

Scholars suggest that media can have a significant impact on shaping cultural norms and values (Bryant & Oliver, 2009). Thus, the role of sexualized media is particularly important in the normalization of sexual harassment and other forms of gender-based violence towards women (Maass et al., 2013; see also the Media-induced Sexual Harassment framework - MiSH; Galdi & Guizzo, 2021). By disseminating the standard that women can be treated as bodies for sexual pleasure, media encourage men to engage in sexual harassment, increase women's acceptance of sexual harassment, and dissuade bystanders from intervening (MiSH, Galdi & Guizzo, 2021).

For example, men and boys exposed to sexualized images of women in the media (e.g., TV, video games) are more likely to engage in gender-harassing behaviours (Galdi et al., 2014), to blame the victim of sexual harassment (e.g., Bernard et al., 2018) and to endorse attitudes that support rape myths, such as the belief that women who are sexually assaulted are responsible for the assault because of the way they dress or behave (e.g., Driesman et al., 2015). On the other hand, women and girls exposed to media sexual objectification tend to exhibit higher rape myth acceptance and tolerance for sexual harassment as well as to underestimate the seriousness of sexual coercion episodes (e.g., Driesman et al., 2015; Reichl et al., 2018). After exposure to sexualization in the media, and related activation of the concept of women as objects, bystanders may also be slower in recognizing sexually harassing incidents and in providing help to fe-

male victims (e.g., Galdi et al., 2017; see also Pacilli et al., 2017 for other results). Sexualized media effects on sexual harassment are thought to be driven by at least three main mechanisms: dehumanization and reduced empathy toward the victim as well as the activation of traditional gender norms and sexism (see Galdi & Guizzo, 2021; Bernard et al., 2020). Overall, the research on sexual objectification and sexual harassment suggests that these two phenomena are closely linked and that efforts to combat one must also address the other.

Conclusions

Sexual objectification is a complex phenomenon (Loughnan & Pacilli, 2014) that targets more women than men, and can have negative effects on both targets and perpetrators. On the one hand, by reducing women to an object of desire, the perpetrator views women as less than humans and worthy of respect. This has dreadful repercussions as sexual objectification is also linked to greater tolerance of sexual harassment and violence. On the other hand, sexual objectification leads women to internalize the message that their worth as a person is tied primarily to their appearance and sexual desirability and to engage in self-objectification and body monitoring. This is tied to a plethora of harmful effects on women's psychological and mental health.

Interventions to combat sexual objectification are therefore crucial for mitigating these effects. This can include raising awareness of sexual objectification through education programs and media literacy intervention, promoting diversity and inclusion in media, and challenging harmful norms and beliefs around gender and sexuality. For example, encouraging critical analysis of media messages can protect young women from some body image repercussions (e.g., McLean et al., 2016). Disseminating criticizing messages against sexualized media may also encourage women to engage in activism (Guizzo et al., 2017) and reduce men's sexually harassing tendencies (Guizzo & Cadinu, 2021). Educating women about self-objectification and strategies to buffer its consequences (e.g., self-compassion) has also been found effective (Liss & Erchull, 2015; Tylka & Augustus-Horvath, 2011). Body positivity images promoting a broader representation of women's bodies can also increase women's body satisfaction (although with some criticism; see Di Michele et al., 2023).

It is interesting to notice that the majority of the strategies aimed at combating sexual objectification proposed so far focuses on women

because they are often the primary victims of this phenomenon (Pecini et al., 2023). However, male gender norms and socialization play an important role in perpetuating sexually objectifying behaviours towards women (Seabrook et al., 2018; see Galdi & Guizzo, 2021 for a review). Thus, promoting progressive values that challenge the traditional norms of hegemonic masculinity might reduce the perpetration of sexual objectification and have a positive impact both on men and women (Pecini et al., 2023).

References

Ainley, V., & Tsakiris, M. (2013). Body conscious? Interoceptive awareness, measured by heartbeat perception, is negatively correlated with self-objectification. *PloS one*, *8*(2), e55568. https://doi.org/10.1371/journal.pone.0055568

American Psychological Association. (2007). *Report of the APA Task Force on the sexualization of girls*. Retrieved from http://www.apa.org/pi/women/programs/girls/report-full.pdf

Archer, D., Iritani, B., Kimes, D. D., & Barrios, M. (1983). Face-ism: Five studies of sex differences in facial prominence. *Journal of Personality and Social Psychology, 45*, 725-735. https://doi.org/10.1037/0022-3514.45.4.725

Aubrey, J. S. (2006). Effects of Sexually Objectifying Media on Self-Objectification and Body Surveillance in Undergraduates: Results of a 2-Year Panel Study. *Journal of Communication, 56*(2), 366–386. https://doi.org/10.1111/j.1460-2466.2006.00024.x

Aubrey, J. S., & Frisby, C. M. (2011). Sexual objectification in music videos: A content analysis comparing gender and genre. *Mass Communication & Society, 14*(4), 475–501. https://doi.org/10.1080/15205436.2010.513468

Baldissarri C., Andrighetto L., & Volpato C. (2022). The longstanding view of workers as objects: antecedents and consequences of working objectification. *European Review of Social Psychology, 33*(1), 81-130. https://doi.org/10.1080/10463283.2021.1956778

Bareket, O., & Shnabel, N. (2020). Domination and Objectification: Men's Motivation for Dominance Over Women Affects Their Tendency to Sexually Objectify Women. *Psychology of Women Quarterly, 44*(1), 28–49. https://doi.org/10.1177/0361684319871913

Bartky, S. L. (1990). *Femininity and domination: Studies in the*

phenomenology of oppression. Psychology Press.

Bernard, P., Cogoni, C., & Carnaghi, A. (2020). The sexualization–objectification link: Sexualization affects the way people see and feel toward others. *Current Directions in Psychological Science, 29*(2), 134-139. https://doi.org/10.1177/0963721419898818

Bernard, P., Gervais, S. J., Allen, J., Campomizzi, S., & Klein, O. (2012). Integrating sexual objectification with object versus person recognition: The sexualized-body-inversion hypothesis. *Psychological Science, 23*, 469-471. http://dx.doi.org/10.1177/0956797611434748

Bernard, P., Gervais, S. J., & Klein, O. (2018). Objectifying objectification: When and why people are cognitively reduced to their parts akin to objects. *European Review of Social Psychology, 29*(1), 82–121. https://doi.org/10.1080/10463283.2018.1471949

Bernard, P., Legrand, S., & Klein, O. (2018). From bodies to blame: Exposure to sexually objectifying media increases tolerance toward sexual harassment. *Psychology of Popular Media Culture, 7*, 99-112. http://dx.doi.org/10.1037/ppm0000114

Bevens, C. L., & Loughnan, S. (2019). Insights into men's sexual aggression toward women: Dehumanization and objectification. *Sex Roles, 81*(11-12), 713-730. https://doi.org/10.1007/s11199-019-01024-0

Bryant, J., & Oliver, M. B. (2009). *Media effects: Advances in theory and research.* New York: Routledge.

Calogero, R. M. (2004). A test of objectification theory: The effect of the male gaze on appearance concerns in college women. *Psychology of Women Quarterly, 28*(1), 16–21. https://doi.org/10.1111/j.1471-6402.2004.00118.x

Calogero, R. M. (2013a). On objects and actions: Situating self-objectification in a system justification context. In Gervais, S. (Ed.) *Objectification and (De) Humanization* (pp. 97-126). Springer New York.

Calogero, R. M. (2013b). Objects don't object: Evidence that self-objectification disrupts women's social activism. *Psychological Science, 24*(3), 312–318. https://doi.org/10.1177/0956797612452574

Calogero, R. M., Tylka, T. L., Siegel, J. A., Pina, A., & Roberts, T.-A. (2021). Smile pretty and watch your back: Personal safety anxiety and vigilance in objectification theory. *Journal of Personality and Social Psychology, 121*(6), 1195–1222. https://doi.org/10.1037/pspi0000344

Carrotte, E. R., Prichard, I., & Lim, M. S. C. (2017). "Fitspiration" on social media: A content analysis of gendered images. *Journal of medical Internet research, 19*(3), e95. https://doi.org/10.2196/jmir.6368

Cikara, M., Eberhardt, J. L., & Fiske, S. T. (2011). From agents to objects: Sexist attitudes and neural responses to sexualized targets. *Journal of Cognitive Neuroscience, 23*, 540-551. http://dx.doi.org/10.1162/jocn.2010.21497

Dall'Ara, E., & Maass, A. (2000). Studying sexual harassment in the laboratory: Are egalitarian women at higher risk? *Sex Roles, 41*, 681-704. https://doi.org/10.1023/A:1018816025988

Daniel, S., & Bridges, S. K. (2010). The drive for muscularity in men: Media influences and objectification theory. *Body Image, 7*(1), 32–38. https://doi.org/10.1016/j.bodyim.2009.08.003

Daniels, E. A., Zurbriggen, E. L., & Monique Ward, L. (2020). Becoming an object: A review of self-objectification in girls. *Body Image, 33,* 278–299. https://doi.org/10.1016/j.bodyim.2020.02.016

Daubenmier, J. J. (2005). The relationship of yoga, body awareness, and body responsiveness to self-objectification and disordered eating. *Psychology of Women Quarterly, 29*(2), 207–219. https://doi.org/10.1111/j.1471-6402.2005.00183.x

Davies, A. E., Burnette, C. B., & Mazzeo, S. E. (2021). Testing a moderated mediation model of objectification theory among Black women in the United States: The role of protective factors. Sex Roles: A Journal of Research, 84(1-2), 91–101. https://doi.org/10.1007/s11199-020-01151-z

de Beauvoir, S. (1989). *The second sex* (H. M. Parshley, Trans.). New York, NY: Vintage. (Original work published 1952)

De Wilde, M., Casini, A., Bernard, P., Wollast, R., Klein, O., & Demoulin, S. (2020). Two preregistered direct replications of "Objects don't object: Evidence that self-objectification disrupts women's social activism". *Psychological Science, 31*(2), 214-223. https://doi.org/10.1177/0956797619896273

Di Michele, D., Guizzo, F., Canale, N., Fasoli, F., Carotta, F., Pollini, A., & Cadinu, M. (2023). # SexyBodyPositive: When Sexualization Does Not Undermine Young Women's Body Image. *International Journal of Environmental Research and Public Health, 20*(2), 991. https://doi.org/10.3390/ijerph20020991

Driesmans, K., Vandenbosch, L., & Eggermont, S. (2015). Playing a videogame with a sexualized female character increases adolescents' rape myth acceptance and tolerance toward sexual harassment. *Games for Health Journal, 4,* 91–94. https://doi.org/10. 1089/g4h.2014.0055.

European Union Agency for Fundamental Rights. (2014). *Violence against women: An EU-wide survey.* Luxembourg: Publications Office of the European Union.

Fredrickson, B. L., & Roberts, T.-A. (1997). Objectification Theory. *Psychology of Women Quarterly*, *21*(2), 173–206. https://doi.org/10.1111/j.1471-6402.1997.tb00108.x

Fredrickson, B. L., Roberts, T.-A., Noll, S. M., Quinn, D. M., & Twenge, J. M. (1998). That swimsuit becomes you: Sex differences in self-objectification, restrained eating, and math performance. *Journal of Personality and Social Psychology, 75*(1), 269–284. https://doi.org/10.1037/0022-3514.75.1.269

Galdi, S., Guizzo, F. (2021) Media-Induced Sexual Harassment: The Routes from Sexually Objectifying Media to Sexual Harassment. *Sex Roles, 84,* 645–669. https://doi.org/10.1007/s11199-020-01196-0

Galdi, S., Maass, A., & Cadinu, M. (2014). Objectifying media their effect on gender role norms and sexual harassment of women. *Psychology of Women Quarterly, 38,* 398-413. https://doi.org/10.1177/0361684313515185

Galdi, S., Maass, A., & Cadinu, M. (2017). Defending the victim of sexual harassment: The influence of civil courage and media exposure. *Psychology of Women Quarterly, 41,* 338-351. https://doi.org/10.1177/0361684317709770

Gapinski, K. D., Brownell, K. D., & LaFrance, M. (2003). Body objectification and "fat talk": Effects on emotion, motivation, and cognitive performance. *Sex Roles: A Journal of Research, 48*(9-10), 377–388. https://doi.org/10.1023/A:1023516209973

Gervais, S. J., & Eagan, S. (2017). Sexual objectification: The common thread connecting myriad forms of sexual violence against women. *American Journal of Orthopsychiatry, 87,* 226-232. http://dx.doi.org/10.1037/ort0000257

Gervais, S. J., Vescio, T. K., Förster, J., Maass, A., & Suitner, C. (2012). Seeing women as objects: The sexual body part recognition bias. *European Journal of Social Psychology, 42,* 743-753. http://dx.doi.org/10.1002/ejsp.1890

Grabe, S., Ward, L. M., & Hyde, J. S. (2008). The role of the media in body image concerns among women: A meta-analysis of experimental and correlational studies. *Psychological Bulletin, 134*(3), 460–476. https://doi.org/10.1037/0033-2909.134.3.460

Gray, K., Knobe, J., Sheskin, M., Bloom, P., & Barrett, L. F. (2011). More than a body: Mind perception and the nature of objectification. *Journal of Personality and Social Psychology, 101,* 1207-1220. http://dx.doi.org/10.1037/a0025883

Greenleaf, C. (2005). Self-objectification among physically active

women. *Sex Roles: A Journal of Research, 52*(1-2), 51–62. https://doi.org/10.1007/s11199-005-1193-8

Guizzo, F., & Cadinu, M. (2021). Women, not objects: Testing a sensitizing web campaign against female sexual objectification to temper sexual harassment and hostile sexism. *Media Psychology, 24*(4), 509-537. https://doi.org/10.1080/15213269.2020.1756338

Guizzo, F., Cadinu, M., Galdi, S., Maass, A., & Latrofa, M. (2017). Objecting to objectification: women's collective action against sexual objectification on television. *Sex Roles, 77*, 352-365. https://doi.org/10.1007/s11199-016-0725-8

Guizzo, F., Canale, N., & Fasoli, F. (2021). Instagram Sexualization: When posts make you feel dissatisfied and wanting to change your body. *Body image, 39*, 62-67. https://doi.org/10.1016/j.bodyim.2021.06.005

Guizzo, F., Di Michele, D., & Cadinu, M. (2024). From sexualized media consumption to salary negotiation: The relation between chronic self-objectification processes and women's negotiation intentions. *Media Psychology.* Advance online publication. https://doi.org/10.1080/15213269.2023.2298683

Hatton, E., & Trautner, M. N. (2011). Equal opportunity objectification? The sexualization of men and women on the cover of *Rolling Stone. Sexuality & Culture, 15*, 256-278. https://doi.org/10.1007/s12119-011-9093-2

Heflick, N. A., & Goldenberg, J. L. (2009). Objectifying Sarah Palin: Evidence that objectification causes women to be perceived as less competent and less fully human. *Journal of Experimental Social Psychology, 45*, 598-601. http://dx.doi.org/10.1016/j.jesp.2009.02.008

Heflick, N. A., & Goldenberg, J. L. (2014). Seeing eye to body: The literal objectification of women. *Current Directions in Psychological Science, 23*, 225-229. http://dx.doi.org/10.1177/0963721414531599

Heflick, N. A., Goldenberg, J. L., Cooper, D. P., & Puvia, E. (2011). From women to objects: Appearance focus, target gender, and perceptions of warmth, morality and competence. *Journal of Experimental Social Psychology, 47*, 572-581. http://dx.doi.org/10.1016/j.jesp.2010.12.020

Hirschman, C., Impett, E. A., & Schooler, D. (2006). Dis/embodied voices: What late-adolescent girls can teach us about objectification and sexuality. *Sexuality Research & Social Policy, 3*, 8-20. https://doi.org/10.1525/srsp.2006.3.4.8

Holland, E., Koval, P., Stratemeyer, M., Thomson, F., & Haslam, N. (2017). Sexual objectification in women's daily lives: A smartphone ecological momentary assessment study. *British Journal of Social Psychology,*

56(2), 314-333. https://doi.org/10.1111/bjso.12152

Impett, E. A., Schooler, D., & Tolman, D. L. (2006). To Be Seen and Not Heard: Femininity Ideology and Adolescent Girls' Sexual Health. *Archives of Sexual Behavior, 35*(2), 131–144. https://doi.org/10.1007/s10508-005-9016-0

Jones, B. A., & Griffiths, K. M. (2015). Self-objectification and depression: An integrative systematic review. Journal of Affective Disorders, 171, 22–32. https://doi.org/10.1016/j.jad.2014.09.011

Kahalon, R., Shnabel, N., & Becker, J. C. (2018). Experimental studies on state self-objectification: A review and an integrative process model. *Frontiers in Psychology, 9,* Article 1268. https://doi.org/10.3389/fpsyg.2018.01268

Kant, I. (1963). *Lectures on ethics.* New York: Harper & Row

Kant, I. (1785). *Groundwork of the Metaphysics of Morals,* Mary Gregor (ed.), Cambridge University Press, 1998.

Karsay, K. (2020). Objectification. In *The International Encyclopedia of Media Psychology* (pp. 1–9). Wiley. https://doi.org/10.1002/9781119011071.iemp0141

Karsay, K., Knoll, J., & Matthes, J. (2018). Sexualizing media use and self-objectification: A meta-analysis. *Psychology of Women Quarterly, 42*(1), 9–28. https://doi.org/10.1177/0361684317743019

Kilpela, L. S., Calogero, R., Wilfred, S. A., Verzijl, C. L., Hale, W. J., & Becker, C. B. (2019). Self-objectification and eating disorder pathology in an ethnically diverse sample of adult women: Cross-sectional and short-term longitudinal associations. Journal of Eating Disorders, 7, Article 45. https://doi.org/10.1186/s40337-019-0273-z

Koval, P., Holland, E., Zyphur, M. J., Stratemeyer, M., Knight, J. M., Bailen, N. H., Thompson, R. J., Roberts, T.-A., & Haslam, N. (2019). How does it feel to be treated like an object? Direct and indirect effects of exposure to sexual objectification on women's emotions in daily life. *Journal of Personality and Social Psychology, 116*(6), 885–898. https://doi.org/10.1037/pspa0000161

Kozee, H. B., Tylka, T. L., Augustus-Horvath, C. L., & Denchik, A. (2007). Development and psychometric evaluation of the Interpersonal Sexual Objectification Scale. Psychology of Women Quarterly, 31(2), 176–189. https://doi.org/10.1111/j.1471-6402.2007.00351.x

Langton, R. (2009). *Sexual Solipsism: Philosophical Essays on Pornography Objectification.* Oxford: Oxford University Press.

Liss, M., & Erchull, M. J. (2015). Not hating what you see: Self-compassion may protect against negative mental health variables connected to

self-objectification in college women. *Body Image, 14,* 5–12. https://doi.org/10.1016/j.bodyim.2015.02.006

Loughnan, S., Haslam, N., Murnane, T., Vaes, J., Reynolds, C., & Suitner, C. (2010). Objectification leads to depersonalization: The denial of mind and moral concern to objectified others. *European Journal of Social Psychology, 40,* 709-717. https://doi.org/10.1002/ejsp.755

Loughnan, S., & Pacilli, M. G. (2014). Seeing (and treating) others as sexual objects: Toward a more complete mapping of sexual objectification. *TPM: Testing, Psychometrics, Methodology in Applied Psychology, 21,* 309-325. https://doi.org/10.4473/TPM21.3.6

Maass, A., Cadinu, M., & Galdi, S. (2013). Motivations and consequences of sexual harassment. In M. K. Ryan & N. R. Branscombe (Eds.), *The SAGE handbook of gender and psychology* (pp. 341-358). London, UK: Sage.

Maass, A., Cadinu, M., Guarnieri, G., & Grasselli, A. (2003). Sexual harassment under social identity threat: The computer harassment paradigm. *Journal of Personality and Social Psychology, 85,* 853-870. https://doi.org/10.1037/0022-3514.85.5.853

MacKinnon, C. (1993). *Only Words.* Cambridge: Harvard University Press.

Marx, K. (1964). *Early writings* (T. B. Bottomore, Trans.). New York: McGraw-Hill. (Original work published 1844)

McKinley, N. M., & Hyde, J. S. (1996). The objectified body consciousness scale: Development and validation. *Psychology of Women Quarterly, 20*(2), 181–215. https://doi.org/10.1111/j.1471-6402.1996.tb00467.x

McLean, S. A., Paxton, S. J., & Wertheim, E. H. (2016). Does media literacy mitigate risk for reduced body satisfaction following exposure to thin-ideal media? *Journal of Youth and Adolescence, 45*(8), 1678–1695. https://doi.org/10.1007/s10964-016-0440-3

Mercurio, A. E., & Landry, L. J. (2008). Self-objectification and well-being: The impact of self-objectification on women's overall sense of self-worth and life satisfaction. *Sex Roles, 58,* 458-466. https://doi.org/10.1007/s11199-007-9357-3

Moradi, B., & Huang, Y.-P. (2008). Objectification theory and psychology of women: A decade of advances and future directions. *Psychology of Women Quarterly, 32*(4), 377–398. https://doi.org/10.1111/j.1471-6402.2008.00452.x

Moradi, B., & Tebbe, E. (2022). A Test of Objectification Theory With Sexual Minority Women. *Psychology of Women Quarterly, 46*(2), 226–240. https://doi.org/10.1177/03616843211052525

Nussbaum, M. (1995). Objectification. *Philosophy and Public Affairs, 24,*

249–291.

Pacilli, M. G., Pagliaro, S., Loughnan, S., Gramazio, S., Spaccatini, F., & Baldry, A. C. (2017). Sexualization reduces helping intentions towards female victims of intimate partner violence through mediation of moral patiency. *British Journal of Social Psychology, 56*(2), 293-313. https://doi.org/10.1111/bjso.12169

Pecini, C., Guizzo, F., Bonache, H., Borges-Castells, N., Morera, M. D., & Vaes, J. (2023). Sexual objectification: advancements and avenues for future research. *Current Opinion in Behavioral Sciences, 50*, 101261. https://doi.org/10.1016/j.cobeha.2023.101261

Pryor, J. B., LaVite, C. M., & Stoller, L. M. (1993). A social psychological analysis of sexual harassment: The person/situation interaction. *Journal of Vocational Behavior, 42*, 68-83. http://dx.doi.org/10.1006/jvbe.1993.1005

Quinn, D. M., Kallen, R. W., & Cathey, C. (2006). Body on My Mind: The Lingering Effect of State Self-objectification. *Sex Roles: A Journal of Research, 55*(11-12), 869–874. https://doi.org/10.1007/s11199-006-9140-x

Reichl, A. J., Ali, J. I., & Uyeda, K. (2018). Latent sexism in print ads increases acceptance of sexual assault. *Sage Open, 8*, 1-11. https://doi.org/10.1177/2158244018769755

Roberts, T.-A., Calogero, R. M., & Gervais, S. J. (2018). Objectification theory: Continuing contributions to feminist psychology. In C. B. Travis, J. W. White, A. Rutherford, W. S. Williams, S. L. Cook, & K. F. Wyche (Eds.), *APA handbook of the psychology of women: History, theory, and battlegrounds* (pp. 249–271). American Psychological Association. https://doi.org/10.1037/0000059-013

Rudman, L. A., & Mescher, K. (2012). Of animals and objects: Men's implicit dehumanization of women and likelihood of sexual aggression. *Personality and Social Psychology Bulletin, 38*(6), 734–746. https://doi.org/10.1177/0146167212436401

Saguy, T., Quinn, D. M., Dovidio, J. F., & Pratto, F. (2010). Interacting like a body: Objectification can lead women to narrow their presence in social interactions. *Psychological Science, 21*, 178-182. https://doi.org/10.1177/0956797609357751

Seabrook, R. C., Ward, L. M., & Giaccardi, S. (2018). Why is fraternity membership associated with sexual assault? Exploring the roles of conformity to masculine norms, pressure to uphold masculinity, and objectification of women. *Psychology of Men & Masculinity, 19*(1), 3–13. https://doi.org/10.1037/men0000076

Slater, A., & Tiggemann, M. (2016). Little girls in a grown up world: Exposure to sexualized media, internalization of sexualization messages, and body image in 6–9 year-old girls. *Body Image, 18,* 19–22. https://doi.org/10.1016/j.bodyim.2016.04.004

Smith, S. L., Choueiti, M., Pieper, K., Yao, K., Case, A., & Choi, A. (2019). *Inequality in 1,200 popular films: Examining portrayals of gender, race/ethnicity, LGBTQ & disability from 2007 to 2018.* USC Annenberg Inclusion Initiative. Retrieved from http://assets.uscannenberg.org/docs/aii-inequality-report-2019-09-03.pdf

Stop Street Harassment. (2018). *The facts behind the #MeToo movements: A national study on sexual harassment and assault.* Reston, VA. Retrieved from http://www.stopstreetharassment.org/wp-content/uploads/2018/01/Full-Report-2018-National-Study-on-Sexual-Harassment-and-Assault.pdf.

Strelan, P., & Hargreaves, D. (2005). Women Who Objectify Other Women: The Vicious Circle of Objectification? *Sex Roles: A Journal of Research, 52*(9-10), 707–712. https://doi.org/10.1007/s11199-005-3737-3

Swim J., Hyers L., Cohen L., Ferguson M. (2001). Everyday sexism: Evidence for its incidence, nature, and psychological impact from three daily diary studies. *Journal of Social Issues*, 57, 31–53. https://doi.org/10.1111/0022-4537.00200

Tiggemann, M. (2013). Objectification theory: Of relevance for eating disorder researchers and clinicians? *Clinical Psychologist, 17*(2), 35–45. https://doi.org/10.1111/cp.12010

Tiggemann, M., & Kuring, J. K. (2004). The role of body objectification in disordered eating and depressed mood. *British Journal of Clinical Psychology, 43*(3), 299–311. https://doi.org/10.1348/0144665031752925

Tiggemann, M., & Williams, E. (2012). The role of self-objectification in disordered eating, depressed mood, and sexual functioning among women: A comprehensive test of objectification theory. Psychology of Women Quarterly, 36(1), 66–75. https://doi.org/10.1177/0361684311420250

Tylka, T. L., & Augustus-Horvath, C. L. (2011). Fighting self-objectification in prevention and intervention contexts. In R. Calogero, S. Tantleff-Dunn, & J. Thompson (Eds.), *Self-objectification in women: Causes, consequences, and counteractions* (pp. 187–214). Washington, DC: American Psychological Association.

Vaes, J., Cogoni, C., & Calcagnì, A. (2020). Resolving the human–object divide in sexual objectification: How we settle the categorization conflict when categorizing objectified and nonobjectified human

targets. *Social Psychological and Personality Science*, *11*(4), 560-569. https://doi.org/10.1177/1948550619875142

Vaes, J., Cristoforetti, G., Ruzzante, D., Cogoni, C., & Mazza, V. (2019). Assessing neural responses towards objectified human targets and objects to identify processes of sexual objectification that go beyond the metaphor. *Scientific Reports*, *9*, 6699. http://dx.doi.org/10.1038/s41598-019-42928-x

Vaes, J., Loughnan, S., & Puvia, E. (2013). The inhuman body: When sexual objectification becomes dehumanizing. In *Humanness and dehumanization* (pp. 186-204). Psychology Press.

Vaes, J., Paladino, P., & Puvia, E. (2011). Are sexualized women complete human beings? Why men and women dehumanize sexually objectified women. *European Journal of Social Psychology, 41*, 774-785. http://dx.doi.org/10.1002/ejsp.824

Vandenbosch, L., & Eggermont, S. (2013). Sexualization of adolescent boys: Media exposure and boys' internalization of appearance ideals, self-objectification, and body surveillance. *Men and Masculinities, 16*(3), 283–306. https://doi.org/10.1177/1097184X13477866

Vandenbosch, L., Vervloessem, D., & Eggermont, S. (2013). "I might get your heart racing in my skin-tight jeans": Sexualization on music entertainment television. *Communication Studies, 64*, 178-194. https://doi.org/10.1080/10510974.2012.755640

Vasquez, E. A., Ball, L., Loughnan, S., & Pina, A. (2018). The object of my aggression: Sexual objectification increases physical aggression toward women. *Aggressive Behavior*, *44*(1), 5-17. https://doi.org/10.1002/ab.21719

Vaughan-Turnbull, C., & Lewis, V. (2015). Body image, objectification, and attitudes toward cosmetic surgery. *Journal of Applied Biobehavioral Research, 20*(4), 179–196. https://doi.org/10.1111/jabr.12035

Vencill, J. A., Tebbe, E. A., & Garos, S. (2015). It's not the size of the boat or the motion of the ocean: The role of self-objectification, appearance anxiety, and depression in female sexual functioning. Psychology of Women Quarterly, 39(4), 471–483. https://doi.org/10.1177/0361684315587703

Ward, L. M. (2016). Media and sexualization: State of empirical research, 1995–2015. *The Journal of Sex Research*, *53*, 560–577. https://doi.org/10.1080/00224499.2016.1142496

Winn, L., & Cornelius, R. (2020). Self-objectification and cognitive performance: A systematic review of the literature. *Frontiers in Psychology, 11,* Article 20. https://doi.org/10.3389/fpsyg.2020.00020

24. Getting Gender in or out of Mind Minding Gender in Language

Elisa Merkel[1] and Janin Roessel[2]
[1]Fresenius University of Applied Sciences, Germany
[2]Leibniz Institute for the German Language, Germany

> "Arguably, grammatical gender is the most widely investigated language feature affecting social cognition." (Maass et al., 2022)

"6 young leaders who are improving the state of the world" was a heading we read online several years ago. When viewing the picture of these leaders of the World Economic Forum (2018), we noticed a surprise to see five females and only one male. Very subtly the English word "leaders" had formed a picture of predominantly males in our mind—although we were aware of the power of words and of gender stereotypes. There are similar stories people share (e.g., Boiler Inclusion Presentation, 2017). Numerous findings show that gender is in our minds when faced with role nouns or descriptions—often with a predominant male bias. Although most nouns in English are not gendered in terms of grammar (it is a so-called natural gender language); men have served as a default in language (e.g., mankind; see Caliskan et al., 2022; Hellinger & Bußmann, 2015). In other languages—called grammatical gender languages—all nouns (and sometimes also associated words) actually have grammatical gender markers (see information box at the end of this chapter). When translating the term "leader" into German, for instance, "Chef" is the masculine word(stem) and "Chefin" with the suffix "-in" the feminine form. Often, the masculine version is used in a generic fashion, with the intention to include all genders. That is why such forms are called masculine generics. However, they rather appear to be false generics as these

designations have been found to promote biased representations, such as thinking predominantly of males (for overviews see e.g., Gabriel et al., 2018; Maass, Suitner, & Merkel, 2014). Gendering in language aimed at promoting more diverse pictures in our minds frequently elicits heated emotions, instilling reactance when people feel something is forced upon them or when existing language and belief systems are challenged (for an overview see Vergoossen et al., 2020; see also Morgenroth & Ryan, 2021). Different gendered forms have also become associated with different political stances or ideological inferences, which may instill motivation as well as reservations (e.g., Bonnin & Coronel, 2021; Gustaffson Sendén et al., 2015; Kotthoff, 2020).

In this chapter, we are not suggesting some imposed "correct" ways of using language but would like to raise awareness for the multifaceted effects of language. Herein, it is human to feel uncertain or overwhelmed by changes in language or new linguistic recommendations, to be reluctant to alter habits. Also human is the need to be acknowledged and addressed. Language can render people invisible and have manifest consequences. It warrants reflecting this impact of language to develop a sensitivity toward what we are communicating and whom we do or do not address—instances, which often accumulate and may create a reality of their own.

Accordingly, we aim at providing an overview of the effects of language on our ways of thinking and constructing our (gendered) worlds. After first insights from a social cognitive perspective, we will regard tangible consequences at an interpersonal level, which for instance relate to stereotypes or the feeling of being included or ignored. We will then present different strategies to make language more gender-inclusive and shed light on more recent findings about gendered language that moves beyond binary understandings of gender. We invite you to take a brief journey along these roads with a focus on the bases, consequences, and new avenues for the research on gendering in language.

From Minds to Societies - How Language Affects Us

Language influences social cognition in manifold ways (see for example, Maass, Suitner, & Merkel, 2014; Wolff & Holmes, 2011). Language may guide our attention, the categories we think with, and activate particular associations and feelings (see Maass et al., 2022). Numerous findings show that language plays a crucial role for the organization and

maintenance of stereotypes (see for instance Maass et al.'s work on the linguistic intergroup bias, Maass, Suitner, & Merkel, 2014, and Maass, Arcuri, & Suitner, 2014, for overviews). Even subtle linguistic variations in gender-markings are apt to shift our constructions of reality; grammatical gender may elicit gender/-ed inferences, even for inanimate objects and concepts where gender assignment appears arbitrary (for an overview, see Maass et al., 2022).

Research has further shown connections between grammatical gender structures (see information box) and gender-linked attitudes. Experimental studies that asked U.S. participants to complete a survey either in English or in one of two grammatical gender languages (French or Spanish), found that sexist attitudes were less expressed in the English condition compared to Spanish and French (Wasserman & Weseley, 2009). Interestingly, such a tendency was also evident among bilinguals. In Estonia, bilinguals of Russian and Estonian reported more liberalized gender-attitudes when interrogated in Estonian (a genderless language) rather than Russian (a gendered language; Pérez & Tavits, 2019). This was found even though the authors reported that Estonians and Russians in Estonia express similar values and political opinions. Correlative findings at the country-level corroborate this picture. Countries with a dominant grammatical gender language compared to countries with dominant natural gender or genderless languages evidenced on average lower economic and social gender equality (Prewitt-Freilino et al., 2012; for similar findings, see also DeFranza et al., 2020; Jakiela & Ozier, 2020; Pérez & Tavits, 2019). This effect persisted even when controlling for the countries' types of political system, religion, and an index of human development (see Prewitt-Freilino et al., 2012). Interestingly, countries with natural gender languages scored higher than those with genderless languages in the overall index of societal gender equality and in women's political empowerment specifically. Prewitt-Freilino and colleagues reason that natural gender (contrary to gendered) languages do not constantly require their users to pay attention to gender, but they may leverage language as a tool to create more symmetric gender representations—a tool that can hardly be used in genderless languages (for male biases in Turkish and Finnish, see, e.g., Renström et al., 2023).

In sum, these findings suggest that grammatical gender marking seems to be related to gendered associations, attitudes, and gender equality, even on a societal level. We will now delve into micro-level, interpersonal effects of various language forms.

What's in a Gendered Word? – Tangible Consequences

The seminal work of Bem and Bem (1973) about job advertisements in English language, showing that the content and wording of job titles "aids and abeds" the discrimination of women, was the starting point for more language-focused research on this topic. McConnell and Fazio (1996) found that job-titles with "man"-suffixes in English (e.g., chairman) were associated with more stereotypically masculine characteristics, such as intelligent or ambitious, than job-titles with "person"-suffixes, as for instance chairperson. Similarly in an experiment in German, a female candidate was seen as less suitable for a high-status job than a male candidate when masculine job titles (e.g., CEOmasc: "Geschäftsführer") as compared to pair forms (CEOfem/CEOmasc: "Geschäftsführerin/Geschäftsführer") were used (Horvath & Sczesny, 2016). This is in line with the "lack-of-fit"-phenomenon[1], which was found for women regarding (stereotypically male) leadership positions more generally (see Heilman & Caleo, 2018). As the former study illustrates, wording can make a difference when it comes to fit perceptions. In other studies, women felt more belonging, intended to invest more in a job, better identified with the position, and expected to stay longer, when pair forms or neutral designations were used in a fictive job interview instead of masculine forms (Stout & Dasgupta, 2011; see Merkel, 2013, for parallel findings in Italian).

For grammatical gender languages, researchers have likewise found that masculine forms commonly elicit a male bias—which may be prevented or reduced by alternative linguistic forms, such as pair forms or neutralizations (for overviews, see Horvath et al., 2016; Sczesny et al., 2016; Stahlberg et al., 2001; Stahlberg et al., 2007; for a recent high-powered replication, see Keith et al., 2022; for pertaining spontaneous reactions in an Implicit Association Test, see Fatfouta & Sczesny, 2023). These findings span various realms, be it work, legal and political or health domains, or the envisioning of musicians and scientists—indicating that gendered words matter.

Intriguingly, the language we use already affects children. In studies with German and Dutch speaking children (Vervecken et al., 2013), pair forms in contrast to masculine forms led to more imaginations of female jobholders, strengthened the believe that also women can succeed in stereotypically male professions, and enhanced girls' interest in these

---

1   According to this phenomenon, stereotypically feminine attributes do not match stereotypically male-typed positions—there is a lack of fit.

occupations (see also Vervecken & Hannover, 2015). Thus, even at an early age, subtle linguistic cues can pave the way for interests and future professional choices.

A recent meta-analysis (Salwender, 2023) on the effects of different gender strategies on various outcomes linked to gender representations covers more than 19,500 participants and over 350 effect sizes. It attests to robust effects of gender-inclusive forms compared to masculine forms—across contexts varying in gender-stereotypicality, for singular and plural forms, and outcomes related to self- and other-perspectives. In sum, the research converges in showing tangible consequences, but how can we concretely implement gendered language to be inclusive?

How to Gender in the 21ˢᵗ Century?

Roughly two or three main strategies for inclusive gendering were identified (for overviews, see Gabriel et al., 2018; Sczesny et al., 2016): One strategy aims at *visibility* by embedding forms that explicitly refer to different genders. This traditionally implied *feminization* but has moved on to be more inclusive, as we will further outline below. A second strategy is *de-gendering* by using neutralizations or avoiding gendered references. A third strategy, which may be aligned with one of the former, is to create *new words* altogether (i.e., neologisms or *neopronouns* specifically). Table 1 presents a (noncomprehensive) overview of these strategies with concrete examples.

Table 1: Overview of gender-inclusive strategies for exemplary languages

Visibility and neologisms	Example	English version/ translation	Implementation
Word pairs	German: Kolleginnen und Kollegen	colleagues	Explicitly naming the feminine and masculine form
Splitting forms (also: slash-forms)	German: Kolleg/-innen	colleagues	Contracted forms combining the masculine word stem and feminine suffix with an intermediate slash
	Swedish: hon/han	she/he	Pair forms combined with a slash instead of "and"
Capital I	German: Mitarbeiter-Innen	staff members	Contracted forms combining the masculine word stem and feminine suffix by capitalizing the first letter of the suffix
Contracted forms with inclusive signs	French: l´étudiant·e	students	Contracted forms inserting an inclusive sign between the masculine word stem and feminine suffix: mid-dot, gender star/asterisk, colon, gender gap; gender pause in speech: a glottal stop [ʔ] is inserted between word stem and suffix
	German: Student*innen, Student:innen, Student_innen; spoken variant: Student[ʔ]innen		
Inclusive suffixes	Italian: amic*, amicə	friend	Substitution of gendered suffixes with a novel inclusive ending: *, ə (pronounced *schwa*), @, x, e
	Spanish: amig@, amigx, amige		
Neologisms	Swedish: hen Italian: loi, lai German: dey	ze	Creation of new words with a generic or inclusive meaning, here: neopronouns
De-gendering			
Gender-neutral words:			Words that do not specify referent gender or words without explicit gender-marking
Nouns	German: die Kundschaft	the clients	
Adjectives	Spanish: atrayente	attractive	
Changes in language use:			Language forms used (non-)traditionally to convey a gender-neutral or inclusive meaning, e g., nominalized participles in German or English they/them/their used as singular pronouns
Nominalizations	German: die Teilnehmenden	the participating [persons]	
Existing word used generically		they (sg.)	
Adoption of English expressions	German: Trainee		
Rephrasing:			Avoiding gendered person references, e.g., with relative clauses or direct speech
Relative clauses	German: die Person, die den Vorsitz übernahm	the person who chaired the meeting	
Direct speech	Italian: Venite per favore insieme!	Please, come together!	

Note. This overview (adapted from Merkel & Menegatti, 2018) offers an orientation but is not exhaustive due to dynamic developments and the plurality of strategies within and across languages.

Neutralization vs. visibility

Different strategies come with different upsides and downsides. Gender-neutral designations, which rely on existing language forms, appear to be relatively accepted (e.g., Adler & Hansen, 2020; Michaux et al., 2021). They do not make gender additionally salient and many gender-neutral designations are easy to apply, typically comprising only one word. However, without salient gender marking, our minds are not prompted to think of gender in a differentiated way, wherefore we likely rely on dominant associations. Men have traditionally served as a default, so that male associations could often be most accessible (see e.g., Caliskan et al., 2022; Stahlberg et al., 2007; Hellinger & Bußmann, 2015). Accordingly, research has frequently found neutral forms to be less effective in reducing male biases (e.g., Gabriel & Gygax, 2008; Jöckel et al., 2021; Lindqvist et al., 2019; but see also Stout & Dasgupta, 2011). Gabriel et al. (2018) therefore recommend using them in contexts that do not already bear markable gender associations. While professions, for instance, vary in the degree of gender stereotypicality and some may be relatively "neutral" (e.g., Miskersky et al., 2014), it appears likely that we will often have biased gender representations in our heads. To move beyond dominant associations, it would be conducive to explicitly mention genders, thereby making them visible. This visibility may come along with a greater likelihood of negative reactions, but it also carries the opportunity to spur reflection and further increase awareness about gender biases (see also Gabriel et al., 2018).

Traditionally, visibility strategies have focused on feminization by adding a feminine form. Pair forms may elicit more balanced representations of females and males (e.g., Horvath et al., 2016, for German and Italian; Körner et al., 2022, for German; Lindqvist et al., 2019, for English and Swedish; Tibblin et al., 2023, for French), whereas variants that almost look like the feminine form (such as capital-I forms in German; see Table 1) tend to elicit stronger effects, which may yield a female bias in representations (e.g., Blake & Klimmt, 2010; Rothmund & Scheele, 2004). Contrary to neutralizations, such male-female forms make the gender dichotomy salient (Gabriel et al., 2018; for the role of order, see Gabriel et al., 2018; Kesebir, 2017). With feminine forms included, further stereotypical female-linked associations may become activated. Feminine job titles were shown to elicit lower status ratings or salary ascriptions (see Gabriel et al., 2018), but not necessarily affect competence (see Horvath & Sczesny, 2016; Horvath et al., 2016), which aligns with increasing beliefs

in competence equality of women and men (Eagly et al., 2020). Interestingly, more negative evaluations of job holders with feminine titles emerged among male rather than female participants (Budziszewska et al., 2014) and more negative responses were found for a gender-related initiative with feminine versus generic masculine references in Poland but not in Austria—with the latter having a longer history of gendering (Formanowicz et al., 2015; see also Formanowicz et al., 2024; for similar findings in Switzerland)[2] . Accordingly, negative side effects appear to be moderated by what is active in people's minds and how adapted people are to gendered language (see also Gabriel et al., 2018).

Going beyond the gender-binary

The feminization approach is recognized as insufficient as it excludes nonbinary people and identities (e.g., Bigler & Leaper, 2015; Körner et al., 2022). We use nonbinary persons as an umbrella term for people who do not or not only identify as being either female or male, while acknowledging that designations differ and are evolving. Misgendering is a source of psychological distress (see e.g., McLemore, 2018; see also Hagen & Galupo, 2014; Knutson et al., 2019), eliciting, for instance, feelings of invalidation, anger, or alienation (Fath & Proudfoot, 2024; Hekanaho, 2020).

Language (use) has shown various possibilities to be more gender-inclusive. The guidelines by the American Psychological Association, for instance, recommend using singular they as a nongendered pronoun (APA, 2020). Whereas singular they relies on an existing word (with a formerly outdated meaning), an example of newly created forms is hen, which was introduced as a third, gender-neutral and inclusive pronoun to complement Swedish she and he (Gustafsson Sendén et al., 2015). Similar pronouns are also emerging in other languages, such as ze in English (see Lindqvist et al., 2019), iel in French, elle in Spanish (Papadopoulos, 2022; Shroy, 2016) or the pronouns loi and lai in Italian (Baiocco et al., 2023; Mirabella et al., 2024). Another strategy relies on inclusive signs, such

2 In many gendered languages, such as German or Polish, feminine forms are usually more complex (with a specific suffix added to the male form or an alternative form that is longer), thereby creating an asymmetry. In Italian, the suffix "-essa" has been used traditionally for female professions but bears a derogatory connotation. Merkel and colleagues (2012) found that a new symmetrically created form (e.g., "avvocata" instead of "avvocatessa" as a female symmetric form of lawyer compared to the male form "avvocato") may shield against the status loss. Both, the symmetricity and the breach with existing forms may contribute to this effect (see also further on regarding the effectiveness of new terms).

as the asterisk "*" as a special character, also termed gender star, with its radiation symbolizing diverse gender identities. The colon evidences increasing usage in German as well. In French, the mid-dot form (e.g., musician·ne·s) may be similarly used as a more inclusive form (Shroy, 2016). In Spanish, the sign -@ emerged as a suffix combining the feminine suffix -a and the masculine suffix -o, thereby still carrying a binary meaning. The suffixes -x and -e emerged as more inclusive variants (Bonnin & Coronel, 2021; see Table 1 for examples; see also Kirey-Sitnikova, 2021, for Russian). Novel implementations and terminology linked to gender-inclusive language evolve dynamically as language users create, try, and negotiate linguistic expressions[3]. It appears that research lags behind these developments.

However, research increasingly addresses more inclusive gendering strategies, and there is evidence for the effectiveness of these forms. The mid-dot form in French raised estimated percentages of women to a similar degree as pair forms compared to masculine forms (related to a text on a professional meeting; Xiao et al., 2022). Studies in Swedish and English (Lindqvist et al., 2019) described a candidate in a recruitment situation, varying how this person was described. If neutral nouns were used, participants predominantly imagined the candidate to be male. This male bias was significantly reduced with pair forms (han/hon, he/she) or newly created pronouns (hen, ze). By contrast, singular they was not effective. Similar to other neutralization attempts, the effectiveness of singular they may depend on what is most accessible in people's minds—such as stereotypes or the typicality of incumbents with different genders (see von der Malsburg et al., 2020; Renström & Klysing, 2024). In a U.S.-study (Keener & Kotvas, 2022), female participants anticipated to be better included and acknowledged, and to have higher work identification and motivation at work if the job advertisement they had read used singular they or word pairs versus masculine pronouns. Male participants were unaffected in these regards, but both male and female participants perceived the job ads to be more sexist given masculine compared to the more gender-inclusive variants[4].

---

3   See, e.g., overviews in the "Nonbinary Wiki": https://nonbinary.wiki/wiki/Gender_neutral_language_in_Spanish or the Gender in Language Project: https://www.genderin-language.com/
4   Merely adding inclusive abbreviations to masculine job titles, such as "lecturer (m/f/d)", as commonly practiced in German job ads, emerged as insufficient to reduce male biases in fast, spontaneous reactions (Fatfouta & Szcesny, 2023).

Regarding strategies employing a typographic sign to increase inclusiveness, experiments in Germany investigated the gender star form (Körner et al., 2022). In a sentence coherence task, which indirectly assesses mental representations, masculine terms elicited a male bias and pair forms induced balanced representations of females and males. The gender star form, visually similar to German feminine forms, induced a female bias, which was smaller than the male bias linked to the masculine forms. It is notable that the presented target nouns (e.g., neighbors, pharmacists, or concert attendees) were assumed to not bear strong gender associations per se. In studies asking for open responses (e.g., listing three athletes or singers), male biases were reduced with the gender star compared to masculine person references, but no female bias emerged (Keith et al., 2022; Kurz & De Mulder, 2023; see also Zacharski & Ferstl 2023).

The effectiveness of new forms such as ze or the gender star may also stem from them being less familiar or more salient—and thereby potent to stimulate reflection (see Diewald & Steinhauer, 2017; Gabriel et al., 2018). Recent work addresses concerns regarding the readability and comprehensibility of such forms. In experiments targeting lexical access in German (Zacharski et al., 2024), reactions to gender star forms did not differ significantly from reactions to feminine and masculine forms in a student sample. Initially slower reaction times in the non-student sample were overcome within the short period of the study, highlighting the role of adaptation with increasing exposure and familiarity (see also Friedrich et al., 2022).

While the findings presented so far offer evidence for the effectiveness of more inclusive gendering, they remain mute as to whether these forms also promote representations or the awareness of nonbinary people (Bradley et al., 2019). In a study in Poland, participants read a text of an unknown person who was described, either, with a special form of passive voice, as typically used by nonbinary persons in Polish, or with feminine or masculine forms (Hansen & Żółtak, 2022). On the one hand, the text with relatively unknown passive voice forms elicited more negative evaluations. This appeared to be attributable to lower comprehensibility and familiarity. On the other hand, participants were more likely to address the person in gender-neutral ways despite an initial male bias, which may indicate a process of reflection. Zacharski and Ferstl (2023) implemented a word-picture matching task in German, also targeting the representation of nonbinary individuals. The gender star appeared to activate inclusive mental representations and as particularly suited

to activate nonbinary representations. Survey experiments in Sweden offer insights regarding attitudinal effects in broader samples (Tavits & Pérez, 2019): Participants had to describe a cartoon-typed figure. When prompted to use female pronouns or the gender-inclusive pronoun *hen*, participants reported more positive attitudes toward women, gays, and transgender individuals later on.

The perspectives of nonbinary people themselves have been addressed predominantly through qualitative approaches (e.g., Moser & Devereux, 2019; Tordoff et al., 2021). Quantitative studies suggest that people who identify as nonbinary prefer neutralizations and inclusive forms, and that being sensitive to chosen pronouns matters (Hekanaho, 2020; Löhr, 2021; Michaux et al., 2021). In a similar vein, recent experiments (Fath & Proudfoot, 2024) demonstrated that acts of omission where nonbinary people do not see their gender identity represented (e.g., in pertaining options in a questionnaire) may function as a social identity threat and elicit negative emotions. An experimental study among U.S. participants complements this picture (Klysing et al., 2022). Nonbinary and trans individuals who read equal employment opportunity statements (EEO) of an organization reported higher organizational attractiveness, anticipated belonging, and trust particularly if these statements explicitly referred to the equality of diverse genders compared to statements referring to binary genders (men and women) or a control group with no EEO. The explicit mention of multiple genders was more effective in these regards than a statement saying that gender is not of relevance (as a de-gendering strategy). However, the latter was also effective in reducing the concern that one's gender would be a disadvantage among gender minority participants. In that study and a parallel Swedish study (Klysing et al., 2022), participants (in general) envisaged higher gender diversity related to gender nonconforming individuals in the organization (picture selection task) after a de-gendered and particularly after a multigendered EEO compared to a binary-gendered or no EEO.

Even though new inclusive gender forms may meet resistance (e.g., Gabriel et al., 2018), particularly initially and when challenging the gender binary (Broussard et al., 2018; Formanowicz & Hansen, 2022; Gustafsson Sendén et al., 2015; Hekanaho, 2020), these findings illustrate their value and encourage openness to new gender forms—which in turn may facilitate adaptation.

Voicing gender in spoken language

Studies on gendering in spoken language point to effects of gendering also in speech. The aforementioned research on job interviews in English, with women evidencing more sense of belonging and higher job-identification with neutral or pair forms (Stout & Dasgupta, 2011) relied on auditory material in one study. In experiments in German (Vervecken et al., 2013), a person giving a speech was perceived as less sexist[5] and more competent when speaking with gendered pair forms compared to masculine forms. Regarding mental representations, masculine plural forms were associated with a male bias in an auditory sentence coherence task in Spanish (Anaya-Ramírez et al., 2022). Experiments employing this paradigm in Germany (Körner et al., 2024) replicated the male bias given masculine forms, whereas a female bias emerged given generically intended feminine forms and the gender pause (i.e., a glottal stop between the masculine word stem and the feminine suffix). However, the gender pause—although identical to the feminine forms except for the glottal stop—elicited a smaller bias. Jöckel et al. (2021) investigated the effectiveness of gendering in speech in a news setting with adults representative of online users in Germany. The gender pause was most effective in increasing female representations compared to masculine variants. Neutralizations and pair forms did not evidence consistent or significant effects. A similar pattern was observed in a second experiment with youth (aged 8 to 12 years), with the gender pause being most effective. Whereas the adult sample judged the ease of comprehension to be lower when they had listened to the gender pause, the youth sample judged all clips to be equally comprehensible[6].

Grammatical gender in speech previously appeared to have less of an impact, presumably due to the fleetingness of grammatical cues in this modality (Gabriel et al., 2018). New forms such as the gender pause might bring along a critical salience in this modality (see Jöckel et al., 2021) to prompt reflection on gender and move beyond accessible gender associations. The gender pause and neutralizations also emerged as the preferred

5 For participants with relatively positive attitudes toward linguistic equality.
6 Previous research on written gendered language suggests that it typically does not compromise intelligibility or memory (see e.g., Beller & Kazazi, 2013; Blake & Klimmt, 2010; Friedrich et al., 2022)—and if it did, people were found to adapt well (see Gabriel et al., 2018). Similarly, research on nonnative accented speech, which differs from a common standard way of speaking, suggests that listeners are able to adapt well (e.g., Baese-Berk et al., 2013; Weber et al., 2014; Witteman et al., 2014), but stereotypes or prejudice may be linked to lower subjective comprehensibility (see, e.g., Roessel et al., 2019).

variants in spoken language by nonbinary people in a survey study in Germany (Michaux et al., 2021). The accompanying salience may, however, also polarize attitudes (Michaux et al., 2021). The gender pause has elicited heated debates since its emergence in public German television news in 2020, but it has also become more widespread (see also Bonnin & Coronel, 2021, for Spanish). Over time, norms may shift and there may be adaptation effects (see Löhr, 2022; Müller-Spitzer & Ochs, 2024; Zacharski et al., 2024). We might consider Sweden as an example, where the introduction of the Swedish pronoun hen had been met with resentment but a few years later attitudes were predominantly positive; time was the strongest predictor for these attitudes (Gustafsson Sendén et al., 2015).

What Can We Learn?

At this point, we may ask whether there is a tradeoff of acceptance and effectiveness when choosing to gender. Neutralizations appear relatively accepted in written and spoken modalities (Adler & Hansen, 2020; Hekanaho, 2020) and constitute a preferred variant by nonbinary individuals (Michaux et al., 2021). However, depending on dominant gender associations, they may be less potent to reduce common male biases, which calls for visibility strategies. We would like to highlight that gender-inclusive language typically is not an either-or strategy. Elegant and comprehensible gendering may comprise different strategies, such as enhancing accessibility in minds at critical points, using language as a spotlight, and employing neutral terms and rephrased descriptions throughout. A recent corpus-based annotation study of German press texts indicates that roughly 10% of the words were (part of) person references and only about 1% of all tokens would need to be adapted for gender-inclusiveness[7] (Müller-Spitzer et al., 2024). While this may differ depending on the context, different guidelines and examples are (increasingly) available to facilitate inclusive gendering[8]. Merely reading or perceiving gender-inclusive language may also increase its usage (see Koeser et al., 2014). Even

7 Texts from the years 2006 to 2020 were included, before the DPA (major resource in the study and largest news agency in Germany) announced to use more gender-neutral language.
8 Examples for German: https://geschicktgendern.de/ or https://www.genderleicht.de/, and a gender-inclusive language toolkit in English: https://www.iapb.org/wp-content/uploads/2022/03/UN-Compact-Gender-Inclusive-Language-Toolkit.pdf, or a blog post on inclusive language linked to the language learning tool Duolingo: https://blog.duolingo.com/gender-neutral-language-and-pronouns/, see also Elmiger (2024).

such strategies, which initially feel unfamiliar and appear as hurdles, may become increasingly automatic and feel natural (see Maass, Suitner, & Merkel, 2014). This also calls for increasing awareness of the consequences of gendered language, gendered expectations, and diverse gender identities (Formanowicz & Hansen, 2022)—in ways that are accessible and bear low thresholds. Updating conceptualizations and representations of gender in language, in our minds, and societies need to go hand in hand (Morgenroth & Ryan, 2021).

Language affects how we construe our world; it may channel our vision or be a spotlight, making visible what we have paid less attention to. Linguistically excluding people may have accumulative negative consequences on their well-being, striving, and decisions in different realms. Whereas masculine terms have been shown to fail generic intentions, rendering female and nonbinary genders rather invisible, new gendered forms have emerged to be more inclusive. Language is constantly evolving; people's needs drive language change, and we outlined increased inclusiveness and creativity in recent years. Openness, perspective taking, and respect may be bonds within this process. Language itself functions as a social bond—and so may the awareness of its social cognitive effects.

A personal note

Our ways crossed with Anne Maass very early, or even at the beginning of our scientific journeys. She accompanied us proximally or distally during our PhDs and beyond to this day—always with an open ear, a warm hug, and a good tea. Anne is an outstanding person and a mentor with passion for language and radiant enthusiasm. Anne has always inspired us and the people around her, truly igniting research. She has been a living example for us that science can be human and diverse. Our starting point was working on gendered language with Anne, wherefore we dedicate this chapter to her. Anne loves and lives inclusiveness as well as thinking out of the box. This is also evident in her being. When it comes to personality (and gender), commonly the Big Two agency and communion are differentiated, but Anne is inclusive of both, merging warmth and competence with much openness, creativity, and brilliance—thereby also uniting people, research strands, and diversity. Thank you for opening and diversifying our minds!

Information Box

> **Gender-Marking in Language**
>
> It is useful to discern how gender is expressed in grammatical structures. We can roughly differentiate languages into three categories: grammatical, natural, and genderless languages (for an overview, see Merkel, 2013; Merkel & Menegatti, 2018; Stahlberg et al., 2007).
>
> In **grammatical gender languages** (e.g., Germanic, Romance, Slavic), all nouns, pronouns, articles, and, to a certain degree, adjectives are assigned a grammatical gender. It often corresponds with targets' biological sex. However, this does not imply that all nouns designate an entity with a biological sex (such as chair, which is feminine in Italian, "la sedia", and masculine in German, "der Stuhl"). In these languages, grammatically masculine designations are often used in a generic fashion (Hellinger & Bußmann, 2015) because gender-labeling is necessary for almost any person references. For instance, the masculine form of "the scientists" (e.g., in Spanish "los científicos") may be used to refer to a group of scientists, even when consisting of persons with different genders.
>
> In **natural gender languages** (e.g., English, Swedish), nouns are not grammatically marked for gender, but pronouns for humans still convey gender information, like s/he or his/her. Lexically gender-marked words such as mankind also exist and have been used in a generic fashion—reflecting that men have traditionally served as a linguistic default (Caliskan et al., 2022; Hellinger & Bußmann, 2015).
>
> In **genderless languages** (e.g., Turkish, Finnish), most nouns can be used for all genders and gender cues are not part of pronouns and adjectives. There is the possibility to add a word with a distinct lexical meaning to the noun, for instance, to clarify if an expression refers to a boy. However, these languages do not necessitate to constantly refer to gender.

References

Adler, A., & Hansen, K. (2020). Studenten, StudentInnen, Studierende? Aktuelle Verwendungspräferenzen bei Personenbezeichnungen.

Themenheft "Sprache Und Geschlecht": Beiträge Zur Gender-Debatte, 47–63. www.gfds.de/publikationen/

Anaya-Ramírez, A., Grinstead, J., Rivera, M. N., Melamed, D., & Reig-Alamillo, A. (2022). The interpretation of Spanish masculine plural NPs: Are they perceived as uniformly masculine or as a mixture of masculine and feminine? Applied Psycholinguistics, 43(6), 1257–1274. https://doi.org/10.1017/S0142716422000352

APA. (2020). Publication manual of the American Psychological Association: The official guide to APA style (7th ed.). American Psychological Association.

Baese-Berk, M. M., Bradlow, A. R., & Wright, B. A. (2013). Accent-independent adaptation to foreign accented speech. The Journal of the Acoustical Society of America, 133, 174–180. https://doi.org/10.1121/1.4789864

Baiocco, R., Rosati, F., & Pistella, J. (2023). Italian proposal for nonbinary and inclusive language: The schwa as a non-gender–specific ending. Journal of Gay & Lesbian Mental Health, 27(3), 248–253. https://doi.org/10.1080/19359705.2023.2183537

Beller, J., & Kazazi, J. (2013). Is there an effect of gender-fair formulations in the German language? Journal of Unsolved Questions, 3(1), 5–8. https://nbn-resolving.org/urn:nbn:de:0168-ssoar-343626

Bem, S. L., & Bem, D. J. (1973). Does sex-biased job advertising "aid and abet" sex discrimination? Journal of Applied Social Psychology, 3(1), 6–18.

Bigler, R. S., & Leaper, C. (2015). Gendered language: Psychological principles, evolving practices, and inclusive policies. Policy Insights from the Behavioral and Brain Sciences, 2(1), 187–194. https://doi.org/10.1177/2372732215600452

Blake, C., & Klimmt, C. (2010). Geschlechtergerechte Formulierungen in Nachrichtentexten. Publizistik, 55(3), 289–304. https://doi.org/10.1007/s11616-010-0093-2

Bonnin, J. E., & Coronel, A. A. (2021). Attitudes toward gender-neutral Spanish: Acceptability and adoptability. Frontiers in Sociology, 6, Article 629616. https://doi.org/10.3389/fsoc.2021.629616

Boroditsky, L., Schmidt, L. A., & Phillips, W. (2003). Sex, syntax, and semantics. In D. Gentner & S. Goldin-Meadow (Eds.), Language in mind: Advances in the study of language and thought (Issue 22, pp. 61–80).

Bradley, E. D., Salkind, J., Moore, A., & Teitsort, S. (2019). Singular 'they' and novel pronouns: gender-neutral, nonbinary, or both? Proceedings

of the Linguistic Society of America, 4(1), 36. https://doi.org/10.3765/plsa.v4i1.4542

Broussard, K. A., Warner, R. H., & Pope, A. R. D. (2018). Too many boxes, or not enough? Preferences for how we ask about gender in cisgender, LGB, and gender-diverse samples. Sex Roles, 78(9–10), 606–624. https://doi.org/10.1007/s11199-017-0823-2

Budziszewska, M., Hansen, K., & Bilewicz, M. (2014). Backlash over gender-fair language: The impact of feminine job titles on men's and women's perception of women. Journal of Language and Social Psychology, 33(6), 681–691.https://doi.org/10.1177/0261927X14544371

Caliskan, A., Ajay, P. P., Charlesworth, T., Wolfe, R., & Banaji, M. R. (2022). Gender bias in word embeddings. Proceedings of the 2022 AAAI/ACM Conference on AI, Ethics, and Society, 156–170. https://doi.org/10.1145/3514094.3534162

DeFranza, D., Mishra, H., & Mishra, A. (2020). How language shapes prejudice against women: An examination across 45 world languages. Journal of Personality and Social Psychology, 119(1), 7–22. https://doi.org/10.1037/pspa0000188.supp

Diewald, G., & Steinhauer, A. (2017). Richtig Gendern: Wie Sie angemessen und verständlich schreiben. Dudenverlag.

Eagly, A. H., Nater, C., Miller, D. I., Kaufmann, M., & Sczesny, S. (2020). Gender stereotypes have changed: A cross-temporal meta-analysis of U.S. public opinion polls from 1946 to 2018. American Psychologist, 75(3), 301–315. https://doi.org/10.1037/amp0000494.supp

Elmiger, D. (2024). Sammlung / Collection / Colección: Leitfäden für geschlechtergerechte/inklusive Sprache. Guides de langue non sexiste / inclusive. Guidelines for non-sexist / inclusive language. Guías para un lenguaje no sexista / inclusivo (Version 3.0). Université de Genève. https://www.unige.ch/lettres/alman/application/files/2417/1075/6345/2024.03_Leitfadensammlung_V_3.pdf

Fatfouta, R., & Sczesny, S. (2023). Unconscious bias in job titles: Implicit associations between four different linguistic forms with women and men. Sex Roles, 89(11), 774–785. https://doi.org/10.1007/s11199-023-01411-8

Fath, S., & Proudfoot, D. (2024). Devaluation by omission: Limited identity options elicit anger and increase identification. Psychological Science, 35(3), 239–249. https://doi.org/10.1177/09567976231223416

Formanowicz, M. M., Cisłak, A., Horvath, L. K., & Sczesny, S. (2015). Capturing socially motivated linguistic change: How the use of gender-fair language affects support for social initiatives in Austria

and Poland. Frontiers in Psychology, 6, Article 1617. https://doi.org/10.3389/fpsyg.2015.01617

Formanowicz, M. M., & Hansen, K. (2022). Subtle linguistic cues affecting gender in(equality). Journal of Language and Social Psychology, 41(2), 127–147. https://doi.org/10.1177/0261927X211035170

Formanowicz, M. M., Hodel, L., & Sczesny, S. (2024). Why using feminine job titles in German is profitable for women: Ascribed linguistic competence enhances prospects of being hired. Journal of Language and Social Psychology, 43(3), 388–398. https://doi.org/10.1177/0261927X231222881

Friedrich, M. C. G, Drößler, V., Oberlehberg, N., & Heise, E. (2021). The influence of the gender asterisk ("Gendersternchen") on comprehensibility and interest. Frontiers in Psychology, 12, Article 760062. https://doi.org/10.3389/fpsyg.2021.760062

Friedrich, M. C. G., Muselick, J., & Heise, E. (2022). Does the use of gender-fair language impair the comprehensibility of video lectures? – An experiment using an authentic video lecture Manipulating Role Nouns in German. Psychology Learning & Teaching. https://doi.org/10.1177/14757257221107348

Gabriel, U., & Gygax, P. (2008). Can societal language amendments change gender representation? The case of Norway. Scandinavian Journal of Psychology, 49(5), 451–457. https://doi.org/10.1111/j.1467-9450.2008.00650.x

Gabriel, U., Gygax, P. M., & Kuhn, E. A. (2018). Neutralising linguistic sexism: Promising but cumbersome? Group Processes and Intergroup Relations, 21(5), 844–858. https://doi.org/10.1177/1368430218771742

Gustafsson Sendén, M., Bäck, E. A., & Lindqvist, A. (2015). Introducing a gender-neutral pronoun in a natural gender language: the influence of time on attitudes and behavior. Frontiers in Psychology, 6, Article 893. https://doi.org/10.3389/fpsyg.2015.00893

Hagen, D. B., & Galupo, M. P. (2014). Trans* individuals´ experiences of gendered language with health care providers: Recommendations for practitioners. International Journal of Transgenderism, 15(1), 16–34.

Hansen, K., Żółtak, K. (2022). Social perception of non-binary individuals. Archives of Sexual Behavior, 51(2027–2035). https://doi.org/10.1007/s10508-021-02234-y

Heilman, M. E., & Caleo, S. (2018). Combatting gender discrimination: A lack of fit framework. Group Processes & Intergroup Relations, 21(5), 725-744. https://doi.org/10.1177/1368430218761587

Hekanaho, L. (2020). Generic and nonbinary pronouns usage, acceptability

and attitudes [Doctoral dissertation]. University of Helsinki. http://urn.fi/URN:ISBN:978-951-51-6832-0

Hellinger, M., & Bußmann, H. (2015). The linguistic representation of women and men. In M. Hellinger & H. Motschenbacher (Eds.), Gender across languages (Vol. 4, pp. 1–25). John Benjamins Publishing Company. https://doi.org/10.1075/impact.36.01hel

Horvath, L. K., Merkel, E. F., Maass, A., & Sczesny, S. (2016). Does gender-fair language pay off? The social perception of professions from a cross-linguistic perspective. Frontiers in Psychology, 6(1), Article 2018. https://doi.org/10.3389/fpsyg.2015.02018

Horvath, L. K., & Sczesny, S. (2016). Reducing women's lack of fit with leadership? Effects of the wording of job advertisements. European Journal of Work and Organizational Psychology, 25(2), 316–328. https://doi.org/https://doi.org/10.1080/1359432X.2015.1067611

Jakiela, P., & Ozier, O. (2020). Gendered language. IZA Discussion Papers (No. 13126). Institute of Labor Economics. https://www.iza.org/publications/dp/13126/gendered-language

Jöckel, S., Dogruel, L., & Bachofer, R. (2021). Wirkung gendersensibler Ansprachen in Anmoderationen bei Erwachsenen und Heranwachsenden. Publizistik, 66(3–4), 441–462. https://doi.org/10.1007/s11616-021-00682-z

Keener, E., & Kotvas, K. (2022). Beyond he and she: does the singular use of "they, them, their" function generically as inclusive pronouns for cisgender men and women? Gender Issues, 1–21. https://doi.org/10.1007/s12147-022-09297-8

Keith, N., Hartwig, K., & Richter, T. (2022). Ladies first or ladies last: Do masculine generics evoke a reduced and later retrieval of female exemplars? Collabra: Psychology, 8(1), Article 32964. https://doi.org/10.1525/collabra.32964

Kesebir, S. (2017). Word order denotes relevance differences: The case of conjoined phrases with lexical gender. Journal of Personality and Social Psychology, 113(2), 262–279. https://doi.org/http://dx.doi.org/10.1037/pspi0000094

Kirey-Sitnikova, Y. (2021). Prospects and challenges of gender neutralization in Russian. Russian Linguistics, 45(2), 143–158. https://doi.org/10.1007/s11185-021-09241-6

Körner, A., Glim, S., & Rummer, R. (2024). Examining the glottal stop as a mark of gender-inclusive language in German. Applied Psycholinguistics, 45(1), 156–179. https://doi.org/10.1017/S0142716424000018

Kotthoff, H. (2020). Gender-Sternchen, Binnen-I oder generisches Maskulinum, … (Akademische) Textstile der Personenreferenz als Registrierungen? Linguistik Online, 103(3), 105–127. https://doi.org/10.13092/lo.103.7181

Klysing, A., Renström, E. A., Gustafsson-Sendén, M., & Lindqvist, A. (2022). Gender diversity in recruitment: Influence of gender trouble on applicant attraction and evaluation. Journal of Applied Social Psychology, 52, 781– 802. https://doi.org/10.1111/jasp.12809

Knutson, D., Koch, J. M., & Goldbach, C. (2019). Recommended terminology, pronouns, and documentation for work with transgender and non-binary populations. Practice Innovations, 4(4), 214–224. https://doi.org/10.1037/pri0000098

Körner, A., Abraham, B., Rummer, R., & Strack, F. (2022). Gender representations elicited by the gender star form. Journal of Language and Social Psychology, 1, 1–19. https://doi.org/10.1177/0261927X221080181

Körner, A., Glim, S., & Rummer, R. (2024). Examining the glottal stop as a mark of gender-inclusive language in German. Applied Psycholinguistics, 45(1), 156–179. https://doi.org/10.1017/S0142716424000018

Kurz, P., & De Mulder, H. (2023). A star is born? The German gender star and its effects on mental representation. Psychology of Language and Communication, 27(1), 384–404. https://doi.org/10.58734/plc-2023-0018

Lindqvist, A., Renström, E. A., & Gustafsson Sendén, M. (2019). Reducing a male bias in language? Establishing the efficiency of three different gender-fair language strategies. Sex Roles, 81(1), 109–117. https://doi.org/10.1007/s11199-018-0974-9

Löhr, R. (2022). Ich denke, es ist sehr wichtig, dass sich so viele Menschen wie möglich repräsentiert fühlen": Gendergerechte Sprache aus der Sicht nicht-binärer Personen. In G. Diewald & D. Nübling (Eds.), Linguistik - Impulse & Tendenzen: Band 95. Genus - Sexus - Gender (pp. 349–379). DE GRUYTER. https://doi.org/10.1515/9783110746396-012

Maass, A., Arcuri, L., & Suitner, C. (2014). Shaping intergroup relations through language. In The Oxford Handbook of Language and Social Psychology (pp. 157–175). Oxford University Press.

Maass, A., Cervone, C., & Ozdemir, I. (2022). Language and social cognition. In Oxford Research Encyclopedia of Psychology. https://doi.org/10.1093/acrefore/9780190236557.013.279

Maass, A., Suitner, C., & Merkel, E. (2014). Does political correctness

make (social) sense? In J. P. Forgas, O. Vinzcze, & J. László (Eds.), Social cognition and communication (pp. 331–343). Psychology Press.

McConnell, A. R., & Fazio, R. H. (1996). Women as men and people: Effects of gender-marked language. Personality and Social Psychology Bulletin, 22(10), 1004–1013. https://doi.org/https://doi.org/10.1177/01461672962210003

McLemore, K. A. (2018). A minority stress perspective on transgender individuals' experiences with misgendering. Stigma and Health, 3(1), 53–64. https://doi.org/10.1037/sah0000070

Merkel, E. (2013). The two faces of gender-fair language [Doctoral dissertation, University of Padua]. Research Padua Archive. http://paduaresearch.cab.unipd.it/6119/

Merkel, E., Maass, A., & Frommelt, L. (2012). Shielding women against status loss: The masculine form and its alternatives in the Italian language. Journal of Language and Social Psychology, 31(3), 311–320. https://doi.org/10.1177/0261927X12446599

Merkel, E., & Menegatti, M. (2018). Language as a means to promote gender-equality. In G. Sáez & I. Valor-Segura (Eds.), Sexism: Past, present and future perspectives. Nova Science Publishers.

Michaux, V., Méndez, J., & Apel, H. (2021). Mündlich gendern? Gerne. Aber wie genau? Ergebnisse einer Akzeptanzuntersuchung zu Formen des Genderns in der Mündlichkeit. Sprachreport, 37(2), 34–41. https://doi.org/10.14618/sr-2-2021-mich

Mirabella, M., Mazzuca, C., De Livio, C., Di Giannantonio, B., Rosati, F., Lorusso, M. M., … & Giovanardi, G. (2024). The role of language in nonbinary identity construction: Gender words matter. Psychology of Sexual Orientation and Gender Diversity. Advance online publication. https://doi.org/10.1037/sgd0000729

Morgenroth, T., & Ryan, M. K. (2021). The effects of gender trouble: An integrative theoretical framework of the perpetuation and disruption of the gender/sex binary. Perspectives on Psychological Science, 16(6), 1113–1142. https://doi.org/10.1177/1745691620902442

Moser, C., & Devereux, M. (2019). Gender neutral pronouns: A modest proposal. International Journal of Transgenderism, 20(2–3), 331–332. https://doi.org/10.1080/15532739.2016.1217446

Müller-Spitzer, C., & Ochs, S. (2024). Shifting social norms as a driving force for linguistic change: Struggles about language and gender in the German Bundestag. ArXiv, arXiv:2402.03887. https://doi.org/10.48550/arXiv.2402.03887

Müller-Spitzer, C., Ochs, S., Koplenig, A., Rüdiger, J. O., & Wolfer, S.

(2024). Less than one percent of words would be affected by gender-inclusive language in German press texts. Humanities and Social Sciences Communications, 11, Article 1343. https://doi.org/10.1057/s41599-024-03769-w

Papadopoulos, B. (2022). A brief history of gender-inclusive Spanish. Deportate, Esuli, Profughe, 40–48.

Pérez, E. O., & Tavits, M. (2019). Language influences public attitudes toward gender equality. Journal of Politics, 81(1), 81–93. https://doi.org/10.1086/700004

Prewitt-Freilino, J. L., Caswell, T. A., & Laakso, E. K. (2012). The gendering of language: A comparison of gender equality in countries with gendered, natural gender, and genderless languages. Sex Roles, 66(3–4), 268–281. https://doi.org/10.1007/s11199-011-0083-5

Renström, E. A., & Klysing, A. (2024). Ideological origins of resistance against gender-inclusive language reforms: Singular they as a de-gendering or multi-gendering strategy. Political Psychology. Advance online publication. https://doi.org/10.1111/pops.13058

Renström, E. A., Lindqvist, A., Akbas, G., Hekanaho, L., & Sendén, M. G. (2023). Are gender-neutral pronouns really neutral? Testing a male bias in the grammatical genderless languages Turkish and Finnish. Journal of Language and Social Psychology, 42(4), 476–487. https://doi.org/10.1177/0261927X221146229

Roessel, J., Schoel, C., Zimmermann, R., & Stahlberg, D. (2019). Shedding new light on the evaluation of accented speakers: Basic mechanisms behind nonnative listeners' evaluations of nonnative accented job candidates. Journal of Language and Social Psychology, 38(1), 3–32. https://doi.org/10.1177/0261927X17747904

Rothmund, J., & Scheele, B. (2004). Personenbezeichnungsmodelle auf dem Prüfstand. Zeitschrift Für Psychologie, 212(1), 40–54. https://doi.org/10.1026/0044-3409.212.1.40

Salwender, M. (2023). Promoting gender equality at different stages of the employee life cycle: New and cumulated findings [Doctoral dissertation, University of Mannheim, Germany]. MADOC. https://nbn-resolving.org/urn:nbn:de:bsz:180-madoc-653279

Sczesny, S., Formanowicz, M., & Moser, F. (2016). Can gender-fair language reduce gender stereotyping and discrimination? Frontiers in Psychology, 7, Article 25. https://doi.org/10.3389/fpsyg.2016.00025

Shroy, A. J. (2016). Innovations in gender-neutral French: Language practices of nonbinary French speakers on Twitter [Unpublished Manuscript]. University of California, Davis.

Stahlberg, D., Braun, F., Irmen, L., & Sczesny, S. (2007). Representation of the sexes in language. In K. Fiedler (Ed.), Social communication (pp. 163–187). Psychology Press.

Stahlberg, D., Sczesny, S., & Braun, F. (2001). Name your favorite musician: Effects of masculine generics and of their alternatives in German. Journal of Language and Social Psychology, 20(4), 464–469.

Stout, J. G., & Dasgupta, N. (2011). When he doesn't mean you: Gender-exclusive language as ostracism. Personality and Social Psychology Bulletin, 37(6), 757–769. https://doi.org/10.1177/0146167211406434

Tavits, M., & Pérez, E. O. (2019). Language influences mass opinion toward gender and LGBT equality. PNAS, 116(34), 16781–16786. https://doi.org/10.1073/pnas.1908156116

Tibblin, J., van de Weijer, J., Granfeldt, J., & Gygax, P. (2023). There are more women in joggeur· euses than in joggeurs: On the effects of gender-fair forms on perceived gender ratios in French role nouns. Journal of French Language Studies, 33(1), 28-51. https://doi.org/10.1017/S0959269522000217

Tordoff, D. M., Haley, S. G., Shook, A., Kantor, A., Crouch, J. M., & Ahrens, K. (2021). "Talk about bodies": Recommendations for using transgender-inclusive language in sex education curricula. Sex Roles, 84(3–4), 152–165. https://doi.org/10.1007/s11199-020-01160-y

Vergoossen, H. P., Renström, E. A., Lindqvist, A., & Gustafsson Sendén, M. (2020). Four dimensions of criticism against gender-fair language. Sex Roles, 83(5–6), 328–337. https://doi.org/10.1007/s11199-019-01108-x

Vervecken, D., & Hannover, B. (2015). Yes I can! Effects of gender fair job descriptions on children's perceptions of job status, job difficulty, and vocational self-efficacy. Social Psychology, 46(2), 76–92. https://doi.org/10.1027/1864-9335/a000229

Vervecken, D., Hannover, B., & Wolter, I. (2013). Changing (s)expectations: How gender fair job descriptions impact children's perceptions and interest regarding traditionally male occupations. Journal of Vocational Behavior, 82(3), 208–220. https://doi.org/10.1016/j.jvb.2013.01.008

von der Malsburg, T., Poppels, T., & Levy, R. P. (2020). Implicit gender bias in linguistic descriptions for expected events: The cases of the 2016 United States and 2017 United Kingdom elections. Psychological Science, 31(2), 115–128. https://doi.org/10.1177/0956797619890619

Wasserman, B. D., & Weseley, A. J. (2009). ¿Qué? Quoi? Do languages with grammatical gender promote sexist attitudes? Sex Roles, 61(9), 634–643. https://doi.org/10.1007/s11199-009-9696-3

Weber, A., Di Betta, A. M., & McQueen, J. M. (2014). Treack or trit: Adaptation to genuine and arbitrary foreign accents by monolingual and bilingual listeners. Journal of Phonetics, 46, 34–51. https://doi.org/10.1016/j.wocn.2014.05.002

Witteman, M. J., Weber, A., & McQueen, J. M. (2014). Tolerance for inconsistency in foreign-accented speech. Psychonomic Bulletin & Review, 21, 512–519. https://doi.org/10.3758/s13423-013-0519-8

Wolff, P., & Holmes, K. J. (2011). Linguistic relativity. Wiley Interdisciplinary Reviews: Cognitive Science, 2(3), 253–265. https://doi.org/10.1002/wcs.104

Xiao, H., Strickland, B., & Peperkamp, S. (2022). How fair is gender-fair language? Insights from gender ratio estimations in French. Journal of Language and Social Psychology, 42(1), 82–106. https://doi.org/10.1177/0261927X221084643

Zacharski, L., & Ferstl, E. C. (2023). Gendered representations of person referents activated by the nonbinary gender star in German: A word-picture matching task. Discourse Processes, 60(4-5), 294–319. https://doi.org/10.1080/0163853X.2023.2199531

Zacharski, L., Kruppa, A., & Ferstl, E. C. (2024). The readability of the nonbinary gender star in German: Evidence from a lexical decision task. Social Psychological Bulletin. Advance online publication. http://doi.org/10.23668/psycharchives.15477

25. Framing Inequality as Privilege Versus Disadvantage: A Double-Edged Sword

Susanne Bruckmüller
Friedrich-Alexander-University Erlangen-Nürnberg, Germany

Inequality between different social groups has traditionally been a topic of social psychology and it is also one of the most pressing issues of our time (Dixson-Declève et al., 2022; United Nations, 2015). Inequality always has (at least) two sides: a group that is in some way better off or advantaged and another group that is worse off or disadvantaged. Both are vital for fully understanding inequality (Nixon, 2019, Phillips et al., 2022). Yet, both public discourse and scientific work pay more attention to disadvantages and discrimination than to advantages and privilege (Gandy & Zhan, 2005; Jun et al., 2022; Malapally et al., 2024; Nixon, 2019; McIntosh, 2012; Phillips et al., 2022). This systematic asymmetry in how we think and talk about inequality raises the question what the implications are of describing and understanding inequality one way versus the other.

Such implications are likely given previous research on *framing*, that is, on the effects of systematic variations in how a specific semantic content is worded. Starting with Kahneman and Tversky's (1979, 1984) seminal work on this topic, research has by now revealed a plethora of ways in which even subtle variations in framing can affect how information is processed, understood, and responded to (see Amsalem & Zoizner, 2022, Keren, 2011, for overviews, including on limits of these effects). Describing inequality as advantage or disadvantage, e.g., as group A having more than group B versus as group B having less than group A, is one example of so-called *equivalency framing*, that is, variations in wording that do not affect the logical meaning of a statement (Scheufele & Iyengar, 2014).

While examining the effects of such subtle variations may at first glance seem like a rather specialized topic of mainly theoretical relevance, previous research has revealed a number of meaningful effects of framing inequality as advantage versus disadvantage (or as privilege versus discrimination – labels vary between studies). The affected variables range from physiological and emotional reactions to intergroup attitudes, legitimacy appraisals, and policy support (e.g., Bruckmüller et al., 2017; Cihangir et al., 2013; Dietze & Craig, 2021; Harth et al., 2008; Powell et al., 2005). Given the enormous societal challenges associated with economic and other inequalities (see Cervone et al., this volume; Dixson-Declève et al., 2022), understanding how people construct and respond to inequality and what role framing plays in this process becomes a question of high practical importance.

Unfortunately, the available research is rather scattered and can appear somewhat contradictory. For example, in different studies, framing racial inequality as White privilege has sometimes decreased (Powell et al., 2005) and sometimes increased racism (Branscombe et al., 2007). To resolve such inconsistencies, this growing field of research awaits theoretical and empirical integration. Accordingly, this chapter has two goals: to draw attention to the importance of these different framings of inequality, and to provide a first attempt at bringing this literature together. I argue that for a meaningful integration, understanding the dynamics of visibility and invisibility is key.

Privilege and Disadvantage in Social Discourse on Inequality

Since Peggy McIntosh (1988) described her "invisible knapsack" of White privilege, that is, the many, often subtle ways in which she was illegitimately advantaged in everyday life because of her Whiteness, awareness of privilege as a key aspect of inequality has been growing. This is evident in news coverage (e.g., Tempesta, 2019), trending Twitter hashtags such as #MyWhitePrivilege, or the development of educational tools to raise awareness for own privileges (e.g., Case et al., 2014; Ehrke et al., 2020; Pickering, 2023). Yet, for most forms of inequality, a focus on disadvantage remains the dominant framing. Content analyses of U.S. mainstream media find that racial inequality is far more often framed as disadvantage for people of color than as advantage for White people (Gandy & Zhan, 2005; Jun et al., 2022). The same is true for scholarly work on racial inequality (Phillips et al., 2022) and for descriptions of

racial inequality by lay people, in lab experiments as well as in social media (Jun et al., 2022; Malapally et al., 2024). Gender inequality is more often framed as female disadvantage than as male advantage (Malapally & Bruckmüller, 2024; Jun et al., 2022) and explanations of gender inequality by laypersons as well as by scholars focus more often on women than on men (Bruckmüller & Braun, 2020; Phillips et al., 2022). Furthermore, when we (Braun et al., 2023) recently examined how different framings of gender inequality are passed on among laypeople, we found that regardless of how we had originally framed a description of gender inequality, participants in a communication chain quickly (namely, by wave 3) returned to predominantly framing inequality as disadvantages for women. For inequality based on social class, the pattern seems less clear. If anything, there seems to be a tendency to frame inequality as rich people's advantage rather than as poor peoples' disadvantage, at least in the U.S. (Jun et al., 2022).

In sum, there is a pervasive pattern of framing inequality as disadvantage rather than as advantage or privilege, gender and racial inequality in particular. This raises two important questions: where such systematic patters of framing come from, and whether and how they matter. These two questions are related, and accordingly, this chapter touches on both of them. Yet, the main focus is on the latter question, that is, on whether and how the framing of inequality affects how people construct and respond to inequality. This is not only interesting from a theoretical point of view, but also practically relevant, for example, to better understand how informational material or political campaigns should be framed to achieve their goals, or to gauge whether and when diversity trainings that focus on privilege awareness (see above) have the intended effects and when they may backfire.

A Double-Edged Sword? Effects of Advantage and Disadvantage Framing

Providing evidence-based recommendations for which framing to use when is difficult for several reasons. First, the available studies vary in many regards, for example, in what kind of inequality is examined, what methods are used, and what correlates, moderators, or dependent variables are examined. Second, the theoretical approaches underlying these experiments vary – ranging from prospect theory (e.g., Ash & Schmierbach, 2013; Valeri & Borgeson, 2007) to social identity theory (e.g., Brans-

combe et al., 2007; Harth et al., 2008) – and many authors do not provide an elaborate theoretical rationale at all (see Malapally et al., 2025). Third and most importantly, the pattern of results produced by this growing field of research seems rather contradictory with regard to the question which framing is more conducive to challenging versus maintaining inequality. In the following, I will mostly summarize experimental research and I will focus on the implications that the observed effects have for challenging versus maintaining inequality. To illustrate these (seemingly) contradictory implications, I use the metaphor of a double-edged sword that can cut either way.

Why Disadvantage Framing May Be the Sharper Edge

A first reason for why disadvantage framing maybe the more efficient tool, if the goal is to reduce inequality, is that disadvantages are easier to recognize than privileges (Phillips & Jun, 2022), making it easier to draw attention to inequality if it is framed as disadvantage. Furthermore, framing inequality as illegitimate (but not legitimate) disadvantage increases sympathy for disadvantaged outgroups among members of advantaged groups (Harth et al., 2008; Schnepf et al., 2023). It can also lead to lower ingroup favoritism and higher willingness to share with the outgroup than framing the same inequality as ingroup advantage (Harth et al., 2008). Several studies have found inequalities to be perceived as more discriminatory and as less legitimate when they were framed as disadvantage rather than as advantage (Bruckmüller et al., 2017; Dietze & Craig, 2021; Schnepf et al., 2023). Relatedly, disadvantage framing can lead to a higher willingness to act against inequality and/or increased support for policies aimed at reducing inequality (Dietze & Craig, 2021; Schnepf et al., 2022; Valeri & Borgeson, 2007).

At the same time, framing inequality as privilege can be threatening to members of advantaged groups (Lowery et al., 2007). This is evident most directly in cardiovascular reactions indicative of threat following advantage rather than disadvantage framing (Dover, 2022) but also in heightened feelings of guilt (Greenaway et al., 2017; Harth et al., 2008; Powell et al., 2005) and more negative views of the self and one's group (Branscombe, 1998). In educational settings, teaching students from advantaged backgrounds about privilege is often met with negative emotions and rejection (Platt, 2013) and students confronted with ingroup educational privilege sometimes respond with educational disengagement (Lowery & Wout, 2010).

As a result of this threat, framing inequality as advantage can cause reactance among members of advantaged groups, for example in the form of downplaying inequality, denying own privilege (Phillips & Lowery, 2015; 2020), or in the form of derogating the disadvantaged outgroup (Branscombe et al., 2007). Another possible defensive reaction is legitimization of ingroup advantage – and legitimate ingroup advantage in turn increases feelings of pride that are associated with higher ingroup favoritism and lower willingness to help the outgroup (Harth et al., 2008). Accordingly, members of privileged groups often prefer to describe illegitimate inequalities as outgroup disadvantage, but legitimate inequalities as ingroup advantage (Dover, 2022).

These preferences already hint at a first problem with disadvantage framing. Not threatening privileged groups by framing inequality in a way that is most palatable to them (even if this framing may be unpleasant for disadvantaged groups, see below) is an inequality in and of itself. There are also empirical reasons why disadvantage framing maybe problematic and why advantage framing may be more conducive to challenging inequality.

Why Advantage Framing May Be the Sharper Edge

First, disadvantage framing can sometimes lead to more negative evaluations of disadvantaged outgroups (Rosette & Koval, 2018), while advantage framing can lead to reactions that are desirable when the goal is to reduce inequality. For example, Powell et al. (2005) found that the collective guilt that their White participants experienced after White advantage (as opposed to Black disadvantage) framing was associated with lower modern racism (see also Greenaway et al., 2017). Diversity trainings that focus on raising awareness for own privileges can improve attitudes towards disadvantaged outgroups (Ehrke et al., 2020), and some studies have found higher support for redistributive policies following ingroup advantage as compared to outgroup disadvantage framing (Lowery et al., 2012; Rosette & Koval, 2018), presumably motivated by attempts to repair the ingroup's moral image. Among participants whose status as members of the advantaged or disadvantaged groups was less clear, framing economic inequality in terms of how much more top earners made compared to the median income earner (rather than vice versa) increased support for redistributive tax policies among conservative U.S. participants, a group that is usually particularly critical of such measures (Chow & Galak, 2012). These findings suggest that even though making own privilege

visible is uncomfortable for members of advantaged groups (Dover, 2022) and can lead to reactance (Phillips & Lowery, 2015; 2020), it can also lead to more productive reactions that improve intergroup relations (Ehrke et al., 2020) or aim directly at reducing (unfair) advantages (Chow & Galak, 2012; Lowery et al., 2012).

Second, while most research on the effects of advantage and disadvantage framing has examined the psychological reactions of members of advantaged groups, the documented effects on members of disadvantaged groups show that disadvantage framing can have rather undesirable consequences for them. First, the overall tendency to frame inequality as disadvantage (see above) fits with the culturally ingrained habit to understand high-status social groups as normative and lower-status groups as deviations from that normative standard that need to be explained (Hegarty & Bruckmüller, 2013; Pratto & Stewart, 2012; Sue, 2006) – a habit that can reinforce harmful stereotypes and negatively affect the self-view of members of disadvantaged groups (Bruckmüller, 2013; Bruckmüller et al., 2012; Hegarty & Pratto, 2001). Following this logic, critical scholars have long argued that chronically framing inequality as disadvantage portrays disadvantaged groups as problematic deviations from normative standards, while at the same time rendering advantages invisible. It thereby simultaneously stigmatizes disadvantaged groups and normalizes advantaged groups' privilege, which ultimately helps to keep systems of inequality in place (Bonilla-Silva, 2003; Case et al., 2014; McIntosh, 1988; 2012; Pratto & Stewart, 2012).

On an empirical level, when Cihangir and colleagues (2012) confronted female participants with exclusion from a task by labelling it "nothing for women" (a disadvantage framing), participants showed cardiovascular reactions indicative of threat, while labelling it as "something for men" (an advantage framing) resulted in cardiovascular reactions indicative of challenge as well as in attempts to disprove the validity of this exclusion. Moreover, Lowery and Wout (2010) found that framing academic inequality as ingroup disadvantage (as opposed to outgroup privilege) caused Black, Latino, and female students to disengage from their own educational achievements.

Moving from effects on participants' self-view to how people understand and respond to inequality as a whole, recent studies also illustrate potential pitfalls of disadvantage framing and potential benefits of advantage framing for challenging inequality. When we (Bruckmüller & Braun, 2020) asked participants to explain why women are underrepresented in

leadership (disadvantage framing) they produced more explanations focusing on women (e.g., their career choices), fewer explanations focused on men (e.g., networking), more suggestions for interventions aimed at "fixing women" (e.g., special trainings), and fewer suggestions aimed at "fixing the system" (e.g., changing problematic workplace cultures; all effects meant in a relative sense). That is, relative to advantage framing, disadvantage framing led to explanations that blamed the disadvantaged group more and to higher support for interventions aimed at helping selected individuals at the expense of support for interventions aimed at long-term systemic changes. We recently replicated these findings in the context of inequalities based on social class (Braun et al., 2024). These findings suggests that advantage framing may be the sharper edge, if we are interested in not only helping disadvantaged individuals, but in truly challenging systems of inequality.

Summary of the Contradictory Findings

Taken together, on the one hand, framing inequality as disadvantage may be a useful tool for challenging inequality. It raises higher awareness of inequality as problematic and has the potential to garner support for interventions aimed at helping disadvantaged people. At the same time, it avoids reactance and defensiveness from members of privileged groups. On the other hand, this comes at a price. Disadvantage framing can be stigmatizing for members of disadvantaged groups and it draws attention away from advantages and systemic issues as important aspects of inequality that also need to be understood and addressed, if we truly want to challenge inequality (see Nixon, 2019). How can we reconcile these contradictions?

Visibility and Invisibility as the Driving Factors

As contradictory as the findings may seem, they can be explained by the same underlying mechanism – one that is rarely made explicit but often implied in the theoretical rationales of the studies summarized above: a phenomenon referred to as *focalism* (Chambers & Windschitl, 2004). In comparative statements, the target (e.g., women in "women earn less than men") draws more attention than the referent (men in the example). As a result, the attributes of the target become salient while attributes of the referent become less salient or even invisible (Tversky, 1977). For

example, hearing that women earn less than men makes salient the negative experiences of women associated with the gender pay gap and one would more likely consider women-centered reasons for it, such as women more often working part-time or having career breaks due to childbirth. Hearing that men earn more than women makes more salient that men have a higher income than women and makes men-centered reasons for the gender pay gap more salient (e.g., men's work being valued more; higher success in salary negotiations). These salient aspects then weigh more heavily in subsequent judgment, while the experiences of the respective other group (men in the first example, women in the second), in contrast, become harder to see and thus tend to be neglected.

Such processes of focalism make understandable the seemingly contradictory effects summarized above. Disadvantage framing puts disadvantaged groups in the foreground. It may thus often be used by members of disadvantaged groups themselves to draw attention to their situation, and it can arouse sympathy and support from others, including members of advantaged groups (Bruckmüller et al., 2017; Dietze & Craig, 2020; Harth et al., 2008). At the same time, such a spotlight on how and why one('s group) is different from another group can also be rather unpleasant, in particular for members of stigmatized groups, who often feel subjected to particular scrutiny (Bruckmüller, 2013; Crosby et al., 2014), and the increased scrutiny that comes with a focus on a disadvantaged group does have the potential to create rejection from others (Rosette & Koval, 2018). This double-edged nature of disadvantage framing also was evident in the set of studies mentioned above, where participants generated explanations of gender inequality in leadership (Bruckmüller & Braun, 2020). Participants produced more explanations focusing on women when inequality was described as women's underrepresentation than when it was described as men's overrepresentation. In line with the notion of visibility as a double-edged sword, some of these explanations seemed suitable to produce sympathy for women, for example, that they have to grapple with unfair expectations or gender stereotypes. Other explanations, however, implicitly or rather explicitly blamed women, for example, by referring to women's career choices or presumable lack of ambition or skill. Equally interesting are the explanations that became less available (i.e., were mentioned less often) following disadvantage framing, namely, explanations focusing on men. These included explanations that portrayed men's higher representation as the outcome of unfair privileges, such as male networking or men favoring other men, but

also explanations suitable to legitimize inequality, such as presumably superior leadership qualities of men.

Turning to effects on members of privileged groups, focalism and the associated (in)visibility of the ingroup can again explain the seemingly contradictory results above. Advantage framing makes the advantages enjoyed by the ingroup more salient. This can result in pride when the advantage is explicitly legitimate (Harth et al., 2008) but more often results in guilt and increased concern with the moral image of the self and the ingroup (e.g., Greenaway et al., 2017; Lowery et al., 2007). This can then lead to defensive reactions such as denial of inequality or outgroup derogation (e.g., Banscombe et al., 2007; Phillips & Lowery, 2015, 2020) but can also motivate members of advantaged groups to take a stance against inequality and/or to support measures that reduce inequality (Bruckmüller & Braun, 2020; Chow & Galak, 2012).

Taken together, depending on whether one focuses on the positive or the negative implications of making disadvantage versus advantage visible versus invisible, one arrives at different conclusions for whether disadvantage or advantage framing is most effective with regard to challenging (versus maintaining) inequality. What recommendations for framing inequality can one deduct from this? In the following, I provide some suggestions, while being well aware that they are rather preliminary. Further theoretical and empirical integration of the literature on framing inequality is needed, and with such integration should come more refined insights on how to best use this double-edged sword. For now, as often, the answer is "it depends", for example, on the type of inequality, the source of the message, and communication goals.

So How to Use This Sword? Some Thoughts on Moderators and Implications

A first conclusion is that which framing is most promising depends on one's particular communication goals. Disadvantage framing seems particularly suitable to raise awareness of inequality as a problem that needs addressing by making visible the negative outcomes it creates for members of disadvantaged groups. Yet, if the goal is to search for the causes of inequality as well as for the most promising interventions, we need advantage framing as well. Understanding both sides of the coin is vital for fully understanding and successfully addressing inequality (Nixon, 2019), and advantage framing makes some otherwise less visible

components more visible (Braun et al., 2024; Bruckmüller & Braun, 2020; Pratto & Stewart, 2012). This can, however, be rather threatening for members of advantaged groups (Dover, 2022; Lowery et al., 2007; Platt, 2013), raising the question what might be the most promising conditions to start a constructive discussion on privilege. The available research already gives us some first clues here.

First, disadvantage framing dominates in social discourse for most, but not all, inequalities. For wealth inequality, Jun et al. (2022) found advantage framing to be (somewhat) more common than disadvantage framing. In addition, in a recent analysis of the discourse on racial inequality on the social media platform Twitter (now called "X"), we found that even though racial inequality is most often framed as Black disadvantage, when tweets use a White privilege framing, they create more discussion (Malapally et al., 2024). Moreover, there were systematic differences in the topics that are covered by tweets framed as Black disadvantage versus as White privilege. The former contained more inequalities that are often understood as differences in the extent to which groups have or receive more of something negative (e.g., police brutality; ballot rejection), while the latter contained more inequalities that are often understood as differences in having or receiving more of a desirable outcome (e.g., vaccinations; college admissions; additional analysis not reported in the published paper, further information from the first author upon request, but see Malapally & Bruckmüller, 2024, for a systematic test). This fits with the very basic principle of presence being easier to perceive than absence (Treisman & Gelande, 1980) and also gives us some clues as to where privilege may be easier to see and to acknowledge and where it may, therefore, be least difficult to start conversations on privilege.

A second factor is who is talking to whom. Messages about ingroup privileges coming from ingroup members may be more credible and cause less reactance than similar messages coming from the outgroup. Greenaway et al. (2017) only found the effects on guilt and attitudes mentioned above when the respective message came from an ingroup member and for several other studies mentioned above, we can safely assume that participants perceived the researchers as belonging to the same privileged ingroup as themselves (e.g., Harth et al., 2008). In a study by Littleford and Jones (2017), students who imagined taking a class on racial inequality by a White or a Black professor acknowledged racial inequality more when a White professor discussed White privilege as opposed to Black disadvantage; Black professors using White privilege framing instead in-

duced more external motivation to respond without prejudice, which is a rather superficial and temporary improvement. Such ingroup-outgroup dynamics may be particularly relevant if one considers that high-identified members of advantaged groups are the ones most likely to react defensively to privilege information (Branscombe et al., 2007; Lowery et al., 2007) and they may also be the ones for whom it matters most whether such a message comes from an ingroup or an outgroup source. This fits not only with the general observation that group members are usually more open to criticism coming from the ingroup than from an outgroup (*the intergroup sensitivity effect*, Hornsey et al., 2002). It also fits with the more specific recommendation that one of the most useful things allies from advantaged groups can do to support disadvantaged outgroups is to discuss inequality in general, and ingroup privilege in particular, with ingroup members (Nixon, 2019; Selvanathan et al., 2020).

A final recommendation based on previous research would be to create circumstances that feel "safe" for advantaged groups before addressing the threatening issue of privilege. Lowery et al.'s (2007) White participants were more willing to admit White privilege following a self-affirmation manipulation. This recommendation is, however, somewhat tricky. While mutual respect and affirmation are probably always good advice for constructive discussions, privileging the experiences and sensitivities of advantaged groups when discussing inequality is also a, albeit rather subtle, form of inequality.

Limitations

Naturally, the analysis above has its limitations. First, a (narrative) review of the literature can only be as good as the original studies and some of the older studies referenced here suffer from insufficient statistical power, meaning that some of the specific effects need to be regarded with caution. Second, I have mostly relied on experimental work and on a rather narrow definition of framing, namely, as different, but by and large logically equivalent ways of describing inequality (as disadvantage or advantage). This leaves out a range of additional work on the framing of inequality, including from other disciplines, that often use a broader framing definition. Examples are framing poverty as an individual problem indicative of individual failure versus as a problem and failure of the system (Rose & Baumgartner, 2013), or describing poverty on an abstract societal level versus via specific stories of individual people affected by it (Iyengar, 1990).

Including this literature is beyond the scope of this chapter. However, I highly encourage such multi-disciplinary integration, not least because many findings tailor nicely with the importance of visibility suggested here. For example, Iyengar's (1990) participants attributed more responsibility for addressing poverty to society following a more abstract framing, and blamed the poor more following exposure to stories about individual poor people. Moreover, when we compare the outcomes studied in experimental work on privilege and disadvantage framing, the parallels to a definition of framing widely used outside of psychology are striking. Specifically, Entman (p. 52) says "to frame is to select some aspects of a perceived reality and make them more salient in a communicating text, in such a way as to promote a particular problem definition, causal interpretation, moral evaluation, and/or treatment recommendation for the item described [emphasis added by removing some italics from the original text]". Translated into the terminology of social psychology and applied to the context of inequality, this means that different framings enhance the salience of disadvantage or of privilege, affecting the perception of inequality, explanations and attributions for it, legitimacy appraisals, as well as support for interventions, all variables for which experimental studies have found effects (see above).

Multi-disciplinary integration may also be particularly valuable as empirical studies on the effects of privilege and disadvantage framing are limited in their number, partially in their methodological quality, and also in their theoretical and empirical integration (with groups of authors that never or hardly ever cite each other, see Malapally et al., 2025). Yet, even with this limited empirical basis it is evident that how we frame inequality matters.

Conclusion

My goal in this chapter was to illustrate that even though chronically framing inequality with a focus on disadvantaged groups may be useful for making inequality visible and to garner well-meaning attention from privileged groups, this framing also has its dangers (such as stigmatizing disadvantaged groups even further). In addition, to truly understand and to effectively challenge inequality, we also need the other edge of the sword, namely, making privilege visible. While this may be difficult, as this edge is sharp and being made aware of one's own privilege can hurt, I hope to have shown that it is worth the effort.

References

Amsalem, E., & Zoizner, A. (2022). Real, but limited: A meta-analytic assessment of framing effects in the political domain. *British Journal of Political Science*, *52*(1), 221–237. https://doi.org/10.1017/S0007123420000253

Ash, E., & Schmierbach, M. (2013). The effects of gain and loss frames on perceptions of racial inequality. *Howard Journal of Communications, 24*(1), 38-56, https://doi.org/10.1080/10646175.2013.748408

Bonilla-Silva, E. (2003). The structure of racism in color-blind, "post-racial" America. *American Behavioral Scientist, 59*(11), 1358-1376. https://doi.org/10.1177/0002764215586826

Branscombe, N. R. (1998). Thinking about one's gender group's privileges or disadvantages: Consequences for well-being in women and men. *British Journal of Social Psychology*, *37*(2), 167-184. https://doi.org/10.1111/j.2044-8309.1998.tb01163.x

Branscombe, N. R., Schmitt, M. T., & Schiffhauer, K. (2007). Racial attitudes in response to thoughts of White privilege. *European Journal of Social Psychology, 37*(2), 203-215. https://doi.org/10.1002/ejsp.348

Braun, M., Schnepf, J., Malapally, A., Reese, G., Martinly, S., & Bruckmüller, S. (2024). *Framing inequality as the rich having more versus the poor having less shapes explanations of inequality and suggestions for interventions.* Manuscript under review.

Braun, M., Martiny, S. E., & Bruckmüller, S. (2023). From serial reproduction to serial communication: Transmission of the focus of comparison in lay communication about gender inequality. *Human Communication Research*, *49*(1), 35-46. https://doi.org/10.1093/hcr/hqac024

Bruckmüller, S. (2013). Singled out as the effect to be explained: Implications for collective self-esteem. *Personality and Social Psychology Bulletin, 39,* 237-249. https://doi.org/10.1177/0146167212471686

Bruckmüller, S., & Braun, M. (2020). One group's advantage or another group's disadvantage? How comparative framing shapes explanations of, and reactions to, workplace gender inequality. *Journal of Language and Social Psychology*, *39*(4), 457–475. https://doi.org/10.1177/0261927X20932631

Bruckmüller, S., Hegarty, P., & Abele, A. E. (2012). Framing gender differences: Linguistic normativity affects perceptions of power and gender stereotypes. *European Journal of Social Psychology, 42*(2), 210–218. https://doi.org/10.1002/ejsp.858

Bruckmüller, S., Reese, G., & Martiny, S. E. (2017). Is higher inequality less

legitimate? Depends on how you frame it! *British Journal of Social Psychology*, *56*(4), 766–781. https://doi.org/10.1111/bjso.12202

Bruckmüller, S., Ryan, M. K., Haslam, S. A., & Peters, K. (2013). Ceilings, cliffs, and labyrinths: Exploring metaphors for workplace gender discrimination. In *The SAGE Handbook of Gender and Psychology* (pp. 450–464). SAGE Publications, Ltd. https://doi.org/10.4135/9781446269930.n27

Case, K. A., Hensley, R., & Anderson, A. (2014). Reflecting on heterosexual and male privilege: Interventions to raise awareness: heterosexual and male privilege. *Journal of Social Issues*, *70*(4), 722–740. https://doi.org/10.1111/josi.12088

Chambers, J. R., & Windschitl, P. D. (2004). Biases in social comparative judgments: the role of nonmotivated factors in above-average and comparative-optimism effects. *Psychological Bulletin*, *130*(5), 813-838. https://doi.org/10.1037/0033-2909.130.5.813

Chow, R. M., & Galak, J. (2012). The effect of inequality frames on Support for redistributive tax policies. *Psychological Science, 23*, 1467–1469. https://doi.org/10.1177/0956797612450035

Cihangir, S., Scheepers, D., Barreto, M., & Ellemers, N. (2013). Responding to gender-based rejection: Objecting against negative and disproving positive intergroup differentiation. *Social Psychological and Personality Science, 4*(2), 151-158. https://doi.org/10.1177/1948550612448195

Crosby, J., King, M., & Savitsky, K. (2014). The minority spotlight effect. *Social Psychological and Personality Science, 5*(7), 743-750. http://doi.org/10.1177/1948550614527625

Dietze, P., & Craig, M. A. (2021). Framing economic inequality and policy as group disadvantages (versus group advantages) spurs support for action. *Nature Human Behaviour, 5*(3), 349–360. https://doi.org/10.1038/s41562-020-00988-4

Dixson-Decléve, S., Gaffney, O., Ghosh, J., Rockström, J., Stoknes, P. E., & Randers, J. (2022). *Earth for All: A Survival Guide for Humanity. A Report to the Club of Rome.* New Society Publishers.

Dover, T. L. (2022). Not all inequalities are created equal: Inequality framing and privilege threat for advantaged groups. *Group Processes and Intergroup Relations, 25*(3), 746-767. https://doi.org/10.1177/13684302211018741

Ehrke, F., Ashoee, A., Steffens, M. C., & Louvet, E. (2020). A brief diversity training: Raising awareness of ingroup privilege to improve attitudes towards disadvantaged outgroups. *International Journal of Psychology, 55*(5), 732-742. https://doi.org/10.1002/ijop.12665

Entman, R. M. (1993). Framing: Toward Clarification of a Fractured Paradigm. *Journal of Communication*, *43*(4), 51–58. https://doi.org/10.1111/j.1460-2466.1993.tb01304.x

Gandy, O. H., & Zhan, L. (2005). Framing comparative risk: A preliminary analysis. *The Howard Journal of Communications*, *16*, 71–86. https://doi.org/10.1080/10646170590948956

Greenaway, K. H., Fisk, K., & Branscombe, N. R. (2017). Context matters: Explicit and implicit reminders of ingroup privilege increase collective guilt among foreigners in a developing country. *Journal of Applied Social Psychology, 47*(12), 677-681. https://doi.org/10.1111/jasp.12482

Harth, N. S., Kessler, T., & Leach, C. W. (2008). Advantaged group's emotional reactions to intergroup inequality: The dynamics of pride, guilt, and sympathy. *Personality and Social Psychology Bulletin*, *34*(1), 115–129. https://doi.org/10.1177/0146167207309193

Hegarty, P., & Bruckmüller, S. (2013). Asymmetric explanations of group differences: Experimental evidence of Foucault's disciplinary power: asymmetric explanations. *Social and Personality Psychology Compass*, *7*(3), 176–186. https://doi.org/10.1111/spc3.12017

Hegarty, P., & Pratto, F. (2001). The effects of social category norms and stereotypes on explanations for intergroup differences. Journal of Personality and Social Psychology, 80(5), 723–735. https://doi.org/10.1037//0022-3514.80.5.723

Hornsey, M. J., Oppes, T., & Svensson, A. (2002). "It's OK if we say it, but you can't": Responses to intergroup and intragroup criticism. *European Journal of Social Psychology*, *32*(3), 293–307. https://doi.org/10.1002/ejsp.90

Iyengar, S. (1990). Framing responsibility for political issues: The case of poverty. *Political Behavior*, *12*(1), 19–40. https://doi.org/10.1007/BF00992330

Jun, S., Chow, R. M., van der Veen, A. M., & Bleich, E. (2022). Chronic frames of social inequality: How mainstream media frame race, gender, and wealth inequality. *Proceedings of the National Academy of Sciences*, *119*(21), e2110712119. https://doi.org/10.1073/pnas.2110712119

Kahneman, D., & A. Tversky, (1979). Prospect Theory: Analysis of decision under risk. *Econometrica, 47*(2), 263-291. https://doi.org/10.2307/1914185

Kahneman, D., & A. Tversky, (1984). Choices, Values, and Frames. *American Psychologist, 39*(4), 341-350. https://doi.org/10.1037/0003-066X.39.4.341

Keren, G. (2011). *Perspectives on framing*. New York: Psychology Press.

Littleford, L. N., & Jones, J. A. (2017). Framing and source effects on White college students' reactions to racial inequity information. *Cultural Diversity and Ethnic Minority Psychology, 23*(1), 143–153. https://doi.org/10.1037/cdp0000102

Lowery, B. S., Chow, R. M., Knowles, E. D., & Unzueta, M. M. (2012). Paying for positive group esteem: How inequity frames affect whites' responses to redistributive policies. *Journal of Personality and Social Psychology, 102*(2), 323–336. https://doi.org/10.1037/a0024598

Lowery, B. S., Knowles, E. D., & Unzueta, M. M. (2007). Framing inequity safely: Whites' motivated perceptions of racial privilege. *Personality and Social Psychology Bulletin, 33*(9), 1237-1250 https://doi.org/10.1177/0146167207303016

Lowery, B. S., & Wout, D. A. (2010). When inequality matters: The effect of inequality frames on academic engagement. *Journal of Personality and Social Psychology, 98*(6), 956–966. https://doi.org/10.1037/a0017926

Malapally, A., Blombach, A., Heinrich, P., Schnepf, J., & Bruckmüller, S. (2024). Unequal tweets: Black disadvantage is (re)retweeted more, but discussed less than White privilege. *Political Communication, 41*(1), 107-128. https://doi.org/10.1080/10584609.2023.2257624

Malapally A., & Bruckmüller S. (2024). *Talking about privilege: Framing inequality as advantage is more likely for inequality in positive than in negative outcomes.* Personality and Social Psychology Bulletin, published online first August 6 2024. http://doi.org/10.1177/01461672241265779

Malapally, A., Methner, N., Braun, M., Wittenborn, S., & Bruckmüller, S. (2025). *Framing inequality as advantage vs. disadvantage: A systematic review of effects and a two-step model to explain them.* Forthcoming manuscript.

McIntosh, P. (1988). *White privilege and male privilege: A personal account of coming to see correspondences through work in women's studies.* Wellesley College Center for Research on Women, Wellesley, MA.

McIntosh, P. (2012). Reflections and future directions for privilege studies. *Journal of Social Issues, 68*(1), 194–206. https://doi.org/10.1111/j.1540-4560.2011.01744.x

Nixon, S. A. (2019). The coin model of privilege and critical allyship: Implications for health. *BMC Public Health, 19*(1), 1637. https://doi.org/10.1186/s12889-019-7884-9

Phillips, L. T., & Jun, S. (2022). Why benefiting from discrimination is less recognized as discrimination. *Journal of Personality and Social Psychology, 122*(5), 825-852. https://doi.org/10.1037/pspi0000298

Phillips, L. T., Jun, S., & Shakeri, A. (2022). Barriers and boosts: Using inequity frames theory to expand understanding of mechanisms of racial and gender inequity. *Academy of Management Annals*, annals.2020.0314. https://doi.org/10.5465/annals.2020.0314

Phillips, L. T., & Lowery, B. S. (2015). The hard-knock life? Whites claim hardships in response to racial inequity. *Journal of Experimental Social Psychology, 61,* 12–18. https://doi.org/10.1016/j.jesp.2015.06.008

Phillips, L. T., & Lowery, B. S. (2020). I ain't no fortunate one: On the motivated denial of class privilege. *Journal of Personality and Social Psychology, 119*(6), 1403–1422. https://doi.org/10.1037/pspi0000240

Pickering, R. M. (2023). Structures of inequity: Teaching privilege and oppression with a tower-building activity, *Teaching of Psychology*, 50(2), 119-124 https://doi.org/10.1177/00986283211056888

Platt, L. F. (2013). Chapter 13: Blazing the trail: Teaching the privileged about privilege. In K. A. Case (Ed.), Deconstructing privilege: Teaching and learning as allies in the classroom (pp. 207– 222). New York, NY: Routledge.

Powell, A. A., Branscombe, N. R., & Schmitt, M. T. (2005). Inequality as ingroup privilege or outgroup disadvantage: The impact of group focus on collective guilt and interracial attitudes. *Personality and Social Psychology Bulletin, 31*(4), 508–521. https://doi.org/10.1177/0146167204271713

Pratto, F., & Stewart, A. L. (2012). Group dominance and the half-blindness of privilege. *Journal of Social Issues, 68*(1), 28–45. https://doi.org/10.1111/j.1540-4560.2011.01734.x

Rose, M., & Baugartner, F. R. (2013) Framing the poor: Media coverage and U.S. poverty policy, 1960–2008. Policy Studies Journal, 41(1), 22-53. https://doi.org/10.1111/psj.12001

Rosette, A. S., & Koval, C. Z. (2018). Framing advantageous inequity with a focus on others: A catalyst for equity restoration. *Journal of Experimental Social Psychology, 76*, 283-289. https://doi.org/10.1016/j.jesp.2018.03.002

Scheufele, D. A., & Iyengar, S. (2014). The state of framing research: A call for new directions. In: K. Kenski & K. H. Jamieson (eds.), *The Oxford Handbook of Political Communication* (pp. 619-632). Oxford: Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199793471.013.47

Schnepf, J., Reese, G., Bruckmüller, S., Braun, M., Rotzinger, J., & Martiny, S. E. (2023). Justice is in the eye of the perceiver: How comparison framing affects our perception of global inequality through social

emotions and justice sensitivity. *Registered Report Peer Community.* Preprint at: PsyArXiv. https://doi.org/10.31234/osf.io/n72cp

Selvanathan, H. P., Lickel, B., & Dasgupta, N. (2020). An integrative framework on the impact of allies: How identity-based needs influence intergroup solidarity and social movements. *European Journal of Social Psychology, 50*(6), 1344-1361. https://doi.org/10.1002/ejsp.2697

Sue, D. W. (2006). The Invisible Whiteness of Being: Whiteness, White Supremacy, White Privilege, and Racism. In M. G. Constantine & D. W. Sue (Eds.), *Addressing racism: Facilitating cultural competence in mental health and educational settings* (pp. 15–30). John Wiley & Sons, Inc.

Tempesta, E. (2019, April 2). "I spat in a cop's face… and all he did was drive me home": Twitter users reveal the 'most outrageous things' they have gotten away with because of their white privilege. *Daily Mail Online.* https://www.dailymail.co.uk/femail/article-6878345/People-reveal-outrageous-things-theyve-gotten-away-white-privilege.html

Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology*, *12*(1), 97–136. https://doi.org/10.1016/0010-0285(80)90005-5

Tversky, A. (1977). Features of similarity. Psychological Review, 84(4), 327–352. https://doi.org/10.1037/0033-295X.84.4.327

United Nations (2015). Transforming Our World: The 2030 Agenda for Sustainable Development. A/RES/70/1.

Valeri, R. & Borgeson, K. (2007). Reframing affirmative action: Examining the impact on White Americans. *Michigan Sociological Review, 21* (Fall 2007), 193-209. https://www.jstor.org/stable/40969132

Acknowledgements

This book is a tribute to Anne Maass, written by her friends, former students, and colleagues from around the world. More than just shaping research questions and methodologies, Anne has touched the hearts of those fortunate enough to collaborate with her, spreading her passion for science in an inspiring and far-reaching way. Following her retirement from the University of Padova in 2022, she embarked on a new academic journey at NYU Abu Dhabi, making this the perfect moment to acknowledge and celebrate her profound impact on the field.

The volume explores two fundamental processes of social cognition. The first focuses on the construction of shared reality, examining how language shapes our perception of the social and cultural world. Through various perspectives, the book illustrates how storytelling and linguistic structures influence our ability to understand emotions, stereotypes, and interpersonal dynamics. The second key theme is the formation of social categories and their impact on intergroup relations. From stereotype creation to the dehumanization of minority groups, it investigates the cognitive and motivational mechanisms that sustain social inequality.

Blending theoretical insights with empirical research, this book not only honors Anne Maass's scientific legacy but also inspires new reflections on the ways we construct, communicate, and transform our social reality.