The Prioritization of Social Stereotypes

Austin A. Baker

Key words: stereotypes, social bias, cognitive architecture, processing fluency

Abstract
Reflecting on the pervasiveness of discrimination throughout human history, it's clear that stereotypes have a strong hold on us. But why? In this paper I argue that stereotypes occupy a privileged position in our cognitive architecture. Unlike other types of information, stereotypes are prioritized in judgment and decision-making. I present empirical work with Jorge Morales and Chaz Firestone in which we demonstrate how prioritized stereotype information can disrupt even basic perceptual judgments about stereotype-incongruent people (Baker, Morales, & Firestone 2019). This disruption causes interactions with stereotype-incongruent people to be experienced as disfluent, subtly motivating us to discriminate against them in virtue of their incongruent social identities.

1. Introduction

Consider the familiar childhood "riddle". A man and his son are in a terrible car accident where the man is killed instantly. His son is taken to the hospital for emergency surgery. However, upon seeing the boy laying on the table the surgeon exclaims, "I can't operate on him, he's my son!" How can this be? People wrestle with this question, guessing that perhaps the boy was adopted or that the surgeon is his stepfather. Consider a similar case: A mother and her son are in a car accident and the mother is killed. When her son is taken to the hospital the nurse exclaims "I can't attend to him, he's my son!"

Of course, the surgeon is the boy's mother, and the nurse is the girl's father.

Belle et al. (2021) found that less than 30% of people (including people that had heard the riddle before) offered the doctor-as-mother or nurse-as-father answers. Moreover, prior exposure to female physicians, liberal-reported political affiliation, and feminist self-identification did not affect participants' likelihood to give stereotype-incongruent answers. This example nicely exemplifies the force of stereotypes. Our stereotypes—even the ones we don't explicitly endorse—influence the assumptions we make about other people, including something as basic as who we see as a doctor and who we see as a nurse.

Stereotype information achieves this sort of cognitive influence because it's *prioritized* over other kinds of information—even more task-relevant information. In this way, stereotypes violate what is taken to be a central principle of cognition: that more relevant information is prioritized over less relevant information in reasoning and decision-making. I discuss this 'Relevance Principle' in section two. In section three I present my own empirical work which suggests that prioritized stereotype information can impair low-level *perceptual* judgments about stereotype-incongruent people. I conclude by discussing a particularly alarming implication of these findings: stereotype prioritization may cause us

to experience even seemingly benign interactions with stereotype-incongruent people as disfluent and uncomfortable. Appreciating the way stereotype congruence effects even basic cognitive processing can help us better understand why people prefer to interact with stereotype-congruent people (e.g., choosing to see male doctors)—these interactions are experienced as especially fluent, which we erroneously interpret as evidence of competence and trustworthiness. In this way stereotype prioritization can shed novel light on subtle but pervasive patterns of discrimination.

Stereotypes fall under the larger umbrella of *social bias*, which is typically broken down into three subcomponents (Eagly & Chaiken 1998; Petty & Wegener 1998; Dovidio et al. 2010): (1) a *prejudice* is an *affective* attitude towards a social group (e.g., not liking a certain group), (2) a *stereotype* is a *cognitive* attitude, which associates a social group with a particular (often negative) characteristic (e.g., believing a certain group of people are lazy), and (3) *discrimination* is a pattern of motivated *behavior* toward a group (e.g., avoiding engagement with a certain group of people).[1] We see stereotypes emerge early in development. For example, by the age of 2 children start acquiring gender role stereotypes—e.g., 'boys grow up to be doctors', 'girls like princesses', etc. (Bauer 1993; Ruble et al. 2006; Wilbourn & Kee 2010). And between the ages of 2 and 6 children start to more rigidly apply patterns of gender stereotypic thinking—e.g., 'he shouldn't be a nurse', 'she can't like trucks', etc. (Trautner et al. 2003).[2]

Stereotypes can thus be understood as complex cognitive schemas made up of implicit and explicit attitudes related to a social group and its members (Hilton & von Hippel 1996; Valian 1998; 2005). The contents of stereotypes vary widely and can include information about the social roles members of a stereotyped group occupy, physical and psychological characteristics of group members, and hierarchical relationships within or between groups—e.g., women are judged to be less likely to succeed in intellectually demanding tasks than similarly capable men (Bian, Leslie, & Cimpian 2018), young black men are perceived to be larger and more physically threatening than similarly sized young white men (Wilson, Hugenberg, & Rule 2018), and disabled people are considered to be less productive in the workplace than similarly productive nondisabled people (Aidan & McCarthy 2014).

The existent literature on stereotypes has focused almost entirely on demonstrating how some particular stereotype (e.g., 'black men are threatening') impacts some particular judgment (e.g., how

[1] A terminological note: when social psychologists claim that prejudices are affective and stereotypes are cognitive, they typically aren't meaning to claim that prejudices are non-conceptual and that stereotypes can't have affective valence. For example, it's been argued that prejudicial attitudes can have conceptual content (Jackson 2020) and that stereotype beliefs can be associated with positive or negative valences (Schröder et al. 2013).

[2] Why do stereotypes emerge so early in development? One might be inclined to assume that young children form stereotypes by tracking statistical regularities in their environments. For example, maybe the children just encounter more male doctors and thus come to stereotype doctors as men. However, the dominant view is social psychology is that stereotypes are—almost entirely—false and in so far as there's a 'kernel of truth' in a stereotype it's very small (Schneider 2004; Jude & Park 1993; for an opposing view see Jussim et al. 2016). It's also interesting that we see the 'men are doctors' stereotype emerges so early in development given that most pediatricians are women (American Academy of Pediatrics, 2019). This means that many children are stereotyping doctors as men while having never actually met a male doctor. Of course, stereotypes have to be learned somewhere, so it's certainly true that the children are tracking *something* (perhaps conversations with caregivers or representations in the media). But there are still big open questions as to where exactly stereotype information is coming from in socialization.

threatening black men are judged to be). However, questions about the cognitive mechanisms that support stereotyping have received less attention. Because of this, a number of important questions relating to the *scope* of stereotyping remain unanswered—e.g., what triggers the activation of stereotypes and what sorts of judgments can be shaped by stereotypes? "Clear scientific progress toward understanding stereotyping and prejudice," Cox and Devine argue, "requires a clear understanding of the cognitive architecture that underlies stereotypes" (2015). The goal of this paper is to make progress on these questions about cognitive architecture. Drawing on empirical and theoretical considerations, I seek to illuminate the relationship between stereotyping and judgment.

2. Relevance and Stereotyping

Fiske and Taylor famously argued that humans are "cognitive misers" (1984). Our minds are computationally limited; however, efficient decision-making requires quick and accurate processing. Because of this, our cognitive architecture is organized to minimize computationally costly operations as much as possible. In this paper I want to focus on a very basic component of miserly cognition: what information is prioritized in reasoning and judgment.

The totality of our impressions, beliefs, concepts, and memories cannot be activated and retrieved at all times (I use the "information" label to refer to these mental entities collectively). At any moment we have at our cognitive disposal many more external inputs (i.e., sights, sounds, smells, etc.) and internal inputs (i.e., beliefs and memories) than we can process and utilize in reasoning and action. Given these limitations, only a subset of our total store of information can be prioritized at any particular time. It's generally thought that *relevance* is the criterion for prioritization such that more relevant information is prioritized over less relevant information. Sperber and Wilson characterize cognitive relevance (2004, my emphasis):

> An input (a sight, a sound, an utterance, a memory) is relevant to an individual when it connects with background information he has available to yield conclusions that matter to him: say, by answering a question he had in mind, improving his knowledge on a certain topic, settling a doubt, confirming a suspicion, or correcting a mistaken impression.

Though relevance is (notoriously) difficult to define, this sketch will be sufficient for the purposes of the present discussion. We can thus think of a piece of information *i* as being relevant to a cognitive task *t* if retrieving and deploying *i* would—given information the agent has—positively contribute to *t* (of course, positive contribution towards *t* given the information the agent has doesn't mean the agent *in fact* successfully performs *t*[3]). We can also talk about relevance in terms of comparisons: one

[3] In this way, the notion of relance being invoked is meant to be psychological and subjective (i.e., the relevance of information depends on what information the agent has) and therefore possible to be undermined by unknown features of the world. Information can be relevant to a task—i.e., can connect to background information the agent has in a way that positively contributes toward task performance—even if the agent is not able to perform the task successfully. If you want to know where your friend is and you are told that he's usually in his office around this time, this information is relevant to your search given what you know even if happens to be the case this afternoon that he's at a doctor's appointment. Learning that your friend is usually in his office narrows down the places he could be and thus positively contributes to the task of finding him, even if you ultimately fail to successfully locate him. Hence, what counts as a

piece of information i_1 is more relevant to a task t than another piece of information i_2 if deploying i_1 would positively contribute to t to a greater degree than deploying i_2. And finally, a piece of information i is *irrelevant* to a task t if i either doesn't positively contribute to t or hinders performance of t.

What does it say about our cognitive architecture that we prioritize relevant information in cognitive processing? In one sense, not having unrestricted access to all our stored information at all times is limiting. Computing over only relevant subsets of information means that we fall short of ideal rationality (see Simon 1957 and Kahneman 2003 on bounded rationality). However, because relevant information helps us execute tasks we care about, prioritizing information according to relevance efficiently facilitates accurate and tractable cognitive processing given our significant computational limitations. In fact, this has been thought to be a defining feature of human cognition, which has proven difficult to successfully emulate within artificial intelligence (see Fodor 1983, 1987, and Dennett 1978 on the Frame Problem).

This idea that relevant information is prioritized in cognitive tasks figures prominently in the literature on perception (Carrasco 2011), priming (Sperber et al., 1979), language (Attardo 2016), face processing (Bublatzky et al. 2017), emotion recognition (Brosch & Sharma 2005), and in-group dynamics (Constable et al. 2019). Sperber and Wilson argue that we see extensive evidence of relevance-based prioritization because our cognitive architecture has evolved to maximize relevance as a matter of efficiency (2004):

> [M]axim[ising] the relevance of the inputs one processes is simply a matter of making the most efficient use of the available processing resources. [… H]umans do have an automatic tendency to maximise relevance, not because we have a choice in the matter – we rarely do – but because of the way our cognitive systems have evolved. As a result of constant selection pressure towards increasing efficiency, the human cognitive system has developed in such a way that our perceptual mechanisms tend automatically to pick out potentially relevant stimuli, our memory retrieval mechanisms tend automatically to activate potentially relevant assumptions, and our inferential mechanisms tend spontaneously to process them in the most productive way.

Thus, we can understand relevance maximization as an evolved strategy to facilitate efficient use of our limited cognitive resources. How does this specifically manifest in our reasoning and decision-making? One place to look is the philosophical literature on fragmentationalism. Many of the phenomena that are cited as motivations for fragmentationalist theories of mind can be understood as involving relevance-based information prioritization (Lewis 1982; Stalnaker 1984; Egan 2008; Bendaña & Mandelbaum forthcoming). Consider an example put forth by David Lewis (1982). Lewis claimed he at one time believed the following: Nassau St. in Princeton runs East to West, the railroad line runs North to South, and Nassau St. and the railroad line are parallel to one another. How could Lewis have held three such obviously contradictory beliefs?

'positive contribution' to a task will depend on what information the agent has and the task itself. But for the sake of this paper, we can set these issues aside. In all the cases I will discuss it will be clear that one piece of information contributes to the agent's task more than another, without further need to commit to a more specific notion of relevance.

Fragmentationalists explain this by appeal to "mental fragments", which are information subsets of the agent's total information set. "Different fragments," Lewis asserts, "come into action in different situations, and the whole system of beliefs never manifest[s] itself all at once" (ibid, 436). Which mental fragment gets prioritized depends on the agent's situational and psychological context—e.g., what they are thinking about, what they are doing, etc. (Bendaña & Mandelbaum forthcoming). Elga and Rayo refer to these contextual factors as the agent's "elicitation condition", a "choice situation for that agent… prompt[ing them] to deploy information for a particular purpose" (forthcoming, 4-5).[4] The terminology of an 'elicitation condition' can help us unpack the notion of relevance a bit further. Elicitation conditions (i.e., choice situations for agents) determine which information is relevant within a given context. Thus, an elicitation condition *e* prompts the activation of some particular mental fragment *m*, where *m* is comprised of information that is prioritized in virtue of relevance to *e*.

In the Lewis case, the elicitation condition of being asked what direction the railroad runs prioritizes a different set of relevant information than the elicitation condition of being asked the relation between Nassau St. and the railroad line. As such, Lewis is only able to recognize that his beliefs are contradictory when he's finally put in an elicitation condition that activates a mental fragment containing all three pieces of information at the same time (e.g., perhaps he's asked to draw a map of Princeton and must label both Nassau St. and the railroad line). Hence, when it comes to judgment and decision-making, the question really shouldn't be "do you have such-and-such piece of information" (though possessing the information is obviously necessary for successful recall) but rather "are you in the right elicitation condition to prioritize such-and-such piece of information?".

Therefore, we can think of an elicitation condition as comprising the situational and psychological features of an agent's choice situation which determines what information is relevant to a decision-making task. We can thus define relevance-based information prioritization as the 'Relevance Principle':

> *Relevance Principle*: An elicitation condition *e* will cause information to be activated such that more relevant information to *e* will generally be prioritized over information which is either less relevant or irrelevant to *e*. That is, if information token i_1 is more relevant to *e* than information token i_2, i_1 will be prioritized over i_2 in reasoning and decision-making.[5]

[4] In fairness to Elga and Rayo, their account is only committed to there being *a* mapping between elicitation conditions and mental fragments such that elicitations conditions are determining which fragments get activated. All their examples suggest that mental fragments are being activated because information within the fragment is *relevant* to agents' elicitation conditions, but they don't directly address this. Fragmentationalist accounts which are more situated in the cognitive science tradition (e.g., Mandelbaum 2016; Quilty-Dunn & Mandelbaum 2018; Bendaña & Mandelbaum forthcoming) more explicitly assume that a fragment is activated because it's relevant to the agent's elicitation condition. However, I mention Elga and Rayo specifically because their term 'elicitation condition' is a helpful shorthand for the psychological and situational aspects of an agent's choice situation which determine which information fragment is activated.

[5] The Relevance Principle here closely resembles Sperber and Wilson's 'Cognitive Principle of Relevance' (1995 and 2004). However, their principle is meant to apply to all (or most) cognitive faculties while I'm focusing on reasoning.

Note that the Relevance Principle is claiming that relevant information is *generally* prioritized. The "generally" qualification is meant to acknowledge that relevance ordering might not always be perfect given our computational limitations and the competing demands of speed and accuracy in cognitive processing. Think about heuristic errors. When I overestimate the prevalence of plane crashes because I remember reading about a plane crash in the news, I am prioritizing less relevant information (i.e., one gruesome plane crash) over more relevant information I have (i.e., my knowledge about the risks involved in different modes of public transit). However, Tversky and Kahneman (1974) argue that we reason according to the availability heuristic because the heuristic *typically* prioritizes the most relevant information—it tends to be the case that the easiest information to bring to mind is the most relevant. Thus, proponents of the Relevance Principle can explain heuristics cases (where the most relevant information isn't being prioritized) in terms of computational limitations preventing our cognitive architecture from perfectly realizing the Relevance Principle.

The Relevance Principle raises a number of theoretical questions that I cannot properly address here. Most significantly: What is the precise definition of 'relevance'? And how can relevance-based prioritization be achieved in a tractable manner? While these two questions are central to the general project of understanding human cognition, the arguments in this paper will not presuppose a particular answer to them. My main goal is to establish that, given minimal assumptions about the notion of relevance, the way stereotype information is activated and deployed in reasoning does not conform to the principle that information gets prioritized in accordance with its relevance.

In the next section, I will argue that stereotypes constitute a privileged class of concepts in that they can swamp out other more relevant information in the prioritization hierarchy. I will present empirical evidence for a systematic prioritization of stereotype information (without regard to relevance), which can't be explained away (like the heuristics cases) by appeal to computational limitations. More specifically, I will be defending the following thesis:

> *Stereotype prioritization*: Stereotypes constitute a privileged information class such that stereotype information i_{ST} can become activated and prioritized at the expense of non-stereotype information $i_{N\text{-}ST}$ by an elicitation condition e, even when $i_{N\text{-}ST}$ is (significantly) more relevant to e than i_{ST}.

In defending stereotype prioritization, I won't be trying to establish that the Relevance Principle is entirely mistaken. My claim is that there exists *at least* one kind of information that is not prioritized in accordance with the Relevance Principle—namely, stereotype information. This is compatible with information prioritization almost always being sensitive to relevance, as the Relevance Principle suggests. Rather than abandoning the Relevance Principle altogether, I propose that we consider the possibility that stereotype information is special with regards to information prioritization (leaving open the possibility of other kinds of information also being special in this way).

Why might stereotypes be prioritized differently from other kinds of information? While I can't give a definitive answer to this question here, I think two possible answers are worthy of mention. On one hand, stereotype information could be prioritized because the contents of stereotypes are *so* strongly reinforced by existent social and cultural power structures (Glaeser 2005; Steele, 2010; Beeghley 2015; Siegel 2017; Munton 2019; Haslanger 2019).[6] If this were correct, then there wouldn't be anything special about stereotype information per se—any sufficiently reinforced piece of information could violate the Relevance Principle. But perhaps most of our concepts are not as strongly reinforced as our stereotypes.[7]

On the other hand, it's been widely argued that many social cognition capacities are innate and that social information—from face perception, to mindreading, to sorting people into in-groups and outgroups—is processed differently than non-social information (Spelke & Kinzler 2007; Hamlin, Wynn, & Bloom 2007). There's further evidence that many of these social capacities are facilitated by processing in brain regions specifically dedicated to social information—for example, the fusiform face area for faces (Kanwisher & Yovel 2006), the right temporoparietal junction for mindreading (Saxe & Kanwisher 2003), and the medial prefrontal cortex for in-group/out-group distinctions (Apps et al. 2018). Thus, perhaps stereotype information is prioritized at the expense of non-social information because social information is processed differently than other types of information.

I find both the reinforcement and nativist explanations for stereotype prioritization quite compelling. However, in this paper I merely aim to demonstrate *that* stereotypes are cognitively prioritized and not explain *why* they are so prioritized.

3. "You're My Doctor?": The Data

Would people prioritize stereotype information even when they had much more relevant and reliable information to utilize? Jorge Morales, Chaz Firestone, and I designed a series of experiments to investigate this question (2019). On each trial we asked participants to judge whether a person in a headshot labeled as a "doctor" or a "nurse" was facing left or right while looking at their picture. Even though we never prompt participants to attend to gender and the headshot subjects' gender was irrelevant to task performance, we wondered if participants' stereotypes about the genders of doctors and nurses would nonetheless affect their orientation judgments. Would stereotype information be prioritized at the expense of task-relevant information?

But what information *is* relevant to the task? Consider what the Relevance Principle should predict. All participants are in the same elicitation condition—perceptually judging if someone labeled as a

[6] For more on how hierarchical power structures reinforce stereotypes see Steele (2010), Glaeser (2005), Beeghley (2015), Siegel (2017), Munton (2019), Haslanger (2019), and Ayala and Beeghly (2020).

[7] Of course, it could actually be that many of our concepts violate the Relevance Principle. However, given its explanatory power and widespread empirical support, we should probably expect that not *all* types of information could so flagrantly violate the Relevance Principle. It seems reasonable to assume that for information to violate the Relevance Principle it has to be special in *some* sense (e.g., be strongly reinforced, distinctively social, etc.).

"doctor" or "nurse" is facing left or right while looking at them. Perceptual information about the orientation of the headshot subject's shoulders is relevant to—and, in fact, wholly sufficient for—successfully making the perceptual orientation judgment (i.e., to determine if someone is facing left or right you just need to looking look at their shoulders, direction of their heads, etc.). However, stereotype information about the gender of doctors and nurses is irrelevant. Recall that being relevant to a task is to positively contribute to it—stereotyping doctors as being men doesn't help you determine what direction a (male or female) doctor is facing. And moreover, even if participants' stereotype information is activated (perhaps seeing a person labeled "doctor" automatically activates gender stereotypes), stereotype information is still less relevant to the perceptual orientation task than the available perceptual information. Therefore, the Relevance Principle straightforwardly predicts that more relevant perceptual information will be prioritized over irrelevant stereotype information.

3.1 Design

We collected 60 standardized images (80x100px) of physicians from a major medical institution, half women and half men. All images had a salient facing direction (left or right, normed in a separate study) that could be manipulated by flipping the image. On each trial, the question "What's the direction of the doctor's/nurse's shoulders?" was shown for three seconds above an empty frame, before a headshot appeared (see Fig. 1 below). Participants then indicated via a keypress whether the shoulders of the headshot subject were facing left or right (1=left; 2=right) while looking at the headshot. After the keypress, the headshot disappeared, and participants were asked to recall the headshot subject's profession. A version of our experiment can be found online at: https://perceptionresearch.org/stereotypes.

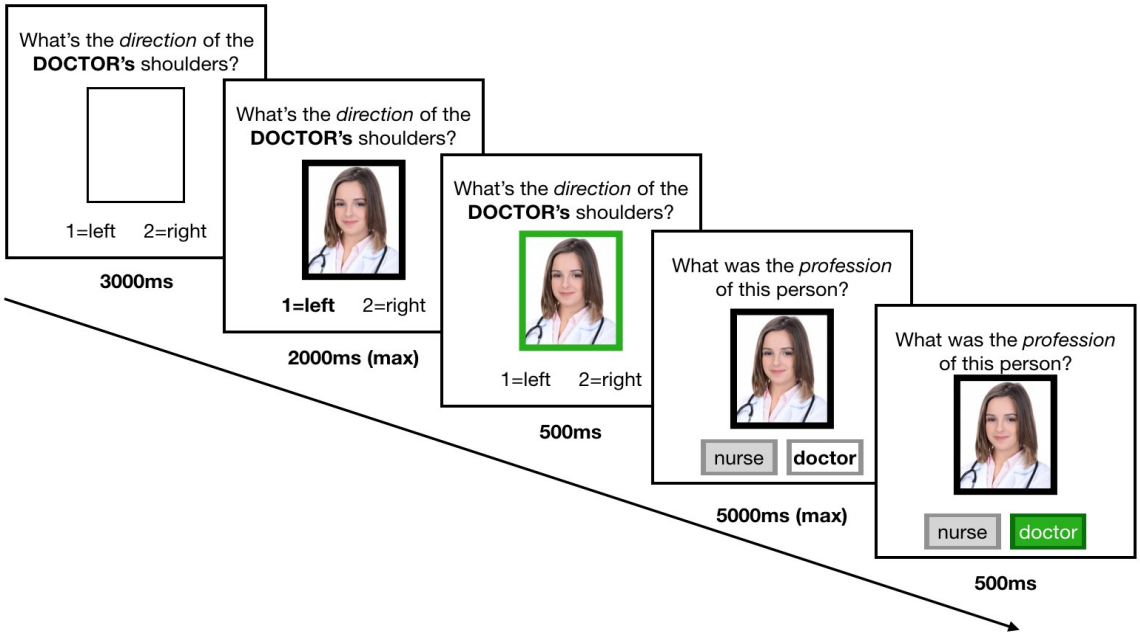

Figure 1. Participants saw the orientation question with the "doctor" or "nurse" profession label for three seconds before the headshot appeared (they could not respond until the image appeared). They then had two seconds to indicate the left or

right orientation of the headshot subject's shoulders via a keypress while looking at the headshot. After they made the keypress, the image disappeared, and they were then asked to remember the headshot subject's profession. Image taken from Baker, Morales, and Firestone (2019) with permission.

Crucially, we introduced a learnable perceptual regularity by manipulating the headshots subjects' orientations: all "doctors" (half of whom were men and half of whom were women) faced one way, and all "nurses" (half men and half women) faced the other way. Learning and applying the perceptual regularity that professional labels predict facing directions makes task-relevant perceptual information *especially* salient—if you see the "doctor" label and expect a left-facing person, you'll be especially sensitive to particularly relevant perceptual features of the headshot subject (e.g., focusing first on the headshot subject's left side, etc.). We made sure that *only* perceptual information was relevant to the task, counterbalancing for gender across "doctors" and "nurses" (i.e., so gender was *never* relevant to headshot subjects' left or right orientation).

[How do we ensure that perceptual information is experienced as more relevant to participants' left/right orientation judgments than stereotype information? Later in this section I'll argue that abstract social information (whether or not it's incidentally activated) is always irrelevant to these sorts of *basic perceptual judgments*. Whether someone's social identity coheres with a previously held stereotype is irrelevant to perceiving basic features of their person like their orientation in space and the color of their clothing (much more on these 'shallow perceptual judgments' in section 3.5). I'll argue that if information prioritization can be accurately captured by the Relevance Principle, it should be the case that relevant perceptual information is always prioritized in *perceptual* judgment over irrelevant non-perceptual information as a matter of efficiency, even if the agent's beliefs are such that they prioritize stereotype information (perhaps irrationally but in coherence with the Relevance Principle) in many other types of judgments. But much more on that later.]

Thus, if as we predicted, irrelevant stereotype information (i.e., doctors are men and nurses are women) is prioritized over task-relevant perceptual information, participants should have difficulty applying the regularity (e.g., doctors face left and nurses face right) when judging the facing direction of stereotype-incongruent people (i.e., female doctors and male nurses). If this is the case, then irrelevant stereotype information is disrupting low-level perceptual processing at the expense of perceptual information. But if only task-relevant perceptual information is prioritized, participants should have no difficulty applying the statistical regularity to both stereotype-congruent and stereotype-incongruent people.

3.2 Statistical Learning

We hypothesized that *statistical learning* might be a possible mechanism for stereotype prioritization. Statistical learning is the process by which we detect patterns and extract meaning from incoming sensory information. It involves attuning to statistical regularities by tracking transitional probabilities—i.e., the frequency of the co-occurrence of two stimuli divided by the frequency of one

stimulus. In other words, perceivers become sensitive to the probability that one stimulus type will accompany or follow another, which allows for the detection of patterns.

Statistical learning is innate (or at least present from birth) and was first discovered in infant language learning. Saffran, Aslin, and Newport (1996) found that when babies are exposed to sequences of syllables, they become sensitive to statistical regularities in syllable order, allowing them to extract units of meaning from the verbal stream and eventually learn words. Statistical learning has also been observed in vision (Vickery et al. 2018), audition (Romberg & Saffran 2010), face perception (Dotsch, Hassin, & Todorov 2017), and social action (Monroy et al. 2017).

We wondered if stereotype information could disrupt the learning and application of basic statistical regularities about stereotype-incongruent people. This is why we introduced the statistically learnable regularity that ("doctor" or "nurse") profession label predicts (left or right) spatial orientation. We hypothesized that participants would have more difficulty applying the regularity to stereotype-incongruent people (i.e., female doctors and male nurses), which would suggest to us that irrelevant stereotype information was being prioritized. Furthermore, discovering that stereotypes could disrupt statistical learning processes would—given the ubiquity of statistical learning—suggest that stereotyping has a much wider scope of potential influence than has been previously appreciated.

3.3 Results

In Experiment 1 (with 100 participants) we found that participants were indeed slower to judge the orientation of stereotype-incongruent headshots (female "doctors" and male "nurses") than stereotype-congruent headshots (male "doctors" and female "nurses"). Experiment 2 directly replicated this result with a larger sample (300 participants) and found similarly robust results.[8]

These findings tell us that stereotype information is being prioritized over task-relevant information in the learning and application of the perceptual regularity even though stereotype information is irrelevant to—and in fact hinders—the orientation task. The prioritization of stereotype information, then, is impairing participants' basic perceptual judgments about stereotype-incongruent people (hereafter, "the prioritization effect"). But can we really be sure that statistical learning is the mechanism of this prioritization? In other words, can we know for sure that the effect is being driven by participants having difficultly applying the learned perceptual regularity (e.g., 'doctors face left') when looking at female doctors because their stereotype that 'doctors are men' is being prioritized? Perhaps the effect is not being driven by statistical learning at all but merely by *brute surprise*—e.g., being told you'll see a nurse and then being surprised three seconds later when you see a stereotype-incongruent man.

[8] In Experiments 1 and 2 the mean reaction times for stereotype-congruent trials and stereotype-incongruent trials were 721ms and 737ms, $t(72)=2.31$, $p<.02$ and 735ms and 750ms, $t(201)=3.80$, $p<.0002$. As part of a larger project, we've been able to replicate this effect in different versions of these experiments with over 900 participants. For more on the methods see Baker, Morales, and Firestone (2019) and see osf.io/6nhrd for our code and raw data.

Thus, there are two possible explanations for the prioritization effect we observe. Stereotypes are impairing participants' perceptual orientation judgments by (1) disrupting the application of the statistically learned regularity or (2) triggering brute surprise. Note that in both explanations, stereotype information is being prioritized at the expense of task-relevant perceptual information. So, whether the mechanism is statistical learning or surprise, participants are prioritizing stereotype information in a way that violates the Relevance Principle. However, in section one I set out to theoretically and empirically explore the cognitive architecture that underlies stereotyping and the relationship between stereotyping and perceptual judgment. Therefore, we ran a version of our experiment to control for the brute surprise possibility, with hopes of establishing a clearer picture of the mechanism underlying prioritization.

To tease apart learning and surprise, in Experiment 3 we repeated Experiment 2 with every aspect of the design held constant except for the statistical regularity between profession and facing direction. If we still observed the prioritization effect in the absence of the perceptual regularity, we'd know the effect doesn't depend on statistical learning and must be driven by surprise. However, if the effect disappeared in the absence of the regularity, we'd be able to conclude that statistical learning is at least necessary for facilitating stereotype prioritization.

In Experiment 3 (with 300 participants) each headshot (regardless of "doctor" or "nurse" profession label) was randomly assigned a facing direction. However, when profession didn't predict orientation, we found that the reaction time difference between stereotype-congruent and stereotype-incongruent trials completely disappears. So, in the absence of the statistically learnable regularity, stereotype information isn't being prioritized. The null result in Experiment 3 gives us a rare glimpse into the cognitive architecture of social stereotypes. Not only do we know that stereotype information in Experiments 1 and 2 is being prioritized over more relevant perceptual information such that basic perceptual judgments about stereotype-incongruent people are impaired; we know this impairment is facilitated by statistical learning processes because in the absence of the perceptual regularity the prioritization effect disappears.

3.4 Explaining the data

Our findings suggest two pieces of information are doing work here. The first, which I will call 'R1' ('R' for 'rule' or 'regularity'), is the gender stereotype—doctors are men and nurses are women. The prioritization effect suggests that contrary to the Relevance Principle, the R1 stereotype *is* activated and prioritized even though R1 isn't relevant to the orientation task. The second, which I will call 'R2', is the perceptual regularity participants statistically learn during the course of our experiment—doctors face one direction (e.g., left) and nurses face the other direction (e.g., right). Again, we know participants implicitly or explicitly learn R2 because the only difference in experimental design between Experiments 1 and 2 and Experiment 3 is the introduction of R2.

> **R1.** (Stereotype) Doctors are men and nurses are women.
> **R2.** (Learned statistical regularity) Doctors face one direction (e.g., left) and nurses face the other direction (e.g., right).

So how does R1 interfere with R2 application? Imagine being one of our participants and seeing the "doctor" label appear above an empty frame. R1 and R2 would motivate the following two expectations about the "doctor" headshot during those three seconds before it appeared: (1) the headshot subject would be a man (because, according to R1, doctors are men) and (2) the headshot subject would be facing left (because, according to R2, doctors face left). In stereotype-congruent "doctor" trials, a left facing man would appear so neither expectation would be violated. However, in stereotype-incongruent "doctor" trials, where participants expected to see a left facing *man* and instead saw a left facing *woman*, participants' expectations would be violated, creating a kind of cognitive lag which would explain the reaction time delay.

Thus, our findings suggest that stereotypes occupy our already limited cognitive resources, crowding out task-relevant information, even when stereotype information is irrelevant to—and impairs—task performance. Our gender stereotypes, then, are so closely intertwined with our doctor and nurse concepts that when we learn about doctors and nurses, we can't help but think we're learning about men and women. Stereotypes are just *that* strong and infectious.

And we shouldn't think that the danger posed by stereotype prioritization is limited to the lab. We've absorbed stereotypes about a wide range of social identities (e.g., gender, race, sexual orientation, disability, etc.), which means that a lot of people we encounter will be 'stereotype-incongruent' in some way. Our findings suggest that it's difficult to learn *anything*—even really *arbitrary* things—about stereotype-incongruent people without stereotype information being prioritized. I'll talk more about the consequences of stereotype prioritization in section four.

3.5 Shallow Perceptual Judgments

In the examples I've given of stereotype information violating the Relevance Principle, irrelevant stereotype information is being prioritized over relevant *perceptual* information such that it is impairing agents' *perceptual* judgments. As I hinted at earlier in the section, I think it matters for my challenge to the Relevance Principle that the judgments participants are making in our experiments are perceptual.

But first, I want to flag a possible objection: because the notion of relevance we're working with is psychological and subjective, how can we be sure that—given the very biased information agents are exposed to—it's really the case that they regard stereotype information as less task-relevant than perceptual information? Afterall, stereotypes are culturally reinforced to such a degree that we might imagine that agents *do* regard stereotype information as especially relevant to all judgments, including the perceptual orientation one we elicit from them. If this were true, then our participants would be prioritizing the most relevant information after all and thus our findings would suggest conformity with the Relevance Principle.

This is where perception becomes significant (it will become clear in a moment why). I want to argue that perceptual information is more relevant than stereotype information because the judgment we elicit from participants is perceptual. But are the orientation judgments perceptual? For one, you might

think the effect we find is not best explained in terms of perceptual *judgment* and but rather in terms of *perception proper*. Perhaps stereotypes are actually altering our perceptions such that stereotype-incongruent female doctors are *seen* as less left facing than male doctors. This should remind readers of the cognitive penetration debate (see Pylyshyn 1999, Firestone & Scholl 2015, and Block 2016). If stereotyping doctors as being men affects how we *perceive* female doctors, then perception isn't entirely encapsulated from cognition. This interpretation is entirely consistent with our findings and would indeed be significant.

However, while our data doesn't rule out cognitive penetration, we can alternatively think about these stereotype prioritization cases in terms of what I call *shallow perceptual judgments*—i.e., immediate judgments that are formed purely on the basis of the shallow outputs of encapsulated perception (Siegel 2017; Quilty-Dunn 2019).[9] These types of judgment are typically thought to be non-inferential and automatic (Bonjour 1985; Lyons 2005; Egan 2008). For example, if the output of perception is an experience as of a rose being red, I will automatically and non-inferentially form a shallow perceptual judgment 'that rose is red'. Shallow perceptual judgments can eventually feed into more complex judgments about, for example, the aesthetic pleasantness of flower bouquets, the likelihood your spouse will think you forgot your wedding anniversary, and even perhaps something as complex as the performativity of modern romance.

Thus, if we model cognition as building towards relative complexity, shallow perceptual judgments can be thought of as the *lowest* level of cognition, which means they should be immune from the influence of more complex cognitive attitudes.[10] In this way, perceptual information will always be most relevant to shallow perceptual judgments, no matter how firmly held and culturally reinforced other cognitive attitudes the agent has might be. Think back to the Relevance Principle. If you're making a high-level judgment about the performativity of modern romance, a vast amount of information could be potentially relevant to that judgment and thus prioritized. However, the *only* information relevant to making the shallow perceptual judgment 'that rose is red' is having the perceptual experience as of there being a red rose. Perhaps I lament that the paradigm rose (as a signifier of romantic love) is red and desire that flower shops sell roses of different colors. Nonetheless, my abstract belief about red roses as a cultural symbol isn't relevant to judging that *this*

[9] This phenomenon comes packaged under different names in the philosophy literature—'perceptual beliefs,' 'perceptual judgments', 'appearance beliefs', etc. I qualify with the 'shallow' label to communicate that I'm not referring to just *any* sort of judgment based on perceptual experience. Specifically, I'm thinking basic judgments that can be roughly mapped to the outputs of perception. I also avoid the 'belief' label, because whether or not shallow perceptual judgments are beliefs will depend on what your account of belief is.

[10] Of course, judging 'that rose is red' builds in object identification and color identification—in judging that it's a red rose you're identifying that it's a rose and that it's red. But it's an open question as to where exactly basic object identification takes place. On one hand, it seems natural to think of object identification as post-perceptual—perception outputs a perceptual experience and then you post-perceptually identify it by applying your category information. However, at least when it comes to *basic* categories, it's been argued that some basic object identification might actually occur *within* the encapsulated perception module (Potter et al. 2014; Mandelbaum 2018). Alternatively still, perhaps rose object identification is post-perceptual but the judgment 'that rose is *red*' requires first making the post-perceptual judgment 'that *is* a rose' such that it would be incorrect to say that 'that rose is red' is at the *lowest* level of cognition. At any rate, if shallow perceptual judgments like 'that rose is red' aren't at the *absolute* lowest level of cognition, these sorts of judgments are still at a very low level and are taking as their input at most perceptual information plus some basic category and color information—no higher-order social or conceptual information.

rose is red. Why would our shallow perceptual judgments (i.e., the lowest level of cognition) be affected by abstract social categories (i.e., very high-level cognition)? If I perceive that a rose is red, then it behooves me to perceptually judge that it's red as quickly and efficiently as possible, regardless of how I feel about red roses as a symbol of romance.

Thinking back to our experiments, even if participants' stereotypes are activated by "doctor" and "nurse" labels, stereotype information is not relevant to the shallow perceptual orientation judgment and thus should not be brought to bear on that judgment at the expense of relevant perceptual information by a relevance-based prioritization mechanism. This will be true for any piece of non-perceptual information and any shallow perceptual judgment. For example, the only information relevant to perceptually judging the color of someone's shirt (or the inflection of their voice, the smell of their breath, etc.) is perceptual information about color (or voice inflection, smell, etc.). That the person is a member of a group you have a strong bias against is irrelevant to your shallow perceptual color judgment. Further, bringing stereotype information to bear in shallow perceptual judgment would just expend limited processing resources that should—if our cognitive architecture is organized around the maximization of relevance—be *exclusively* used to (quickly and accurately) process relevant perceptual information. Irrelevant non-perceptual information (e.g., stereotype information) would just tax our processing resources such that shallow perceptual judgments would be needlessly slowed down (and unlike abstract judgments about romantic performativity, efficient action and decision-making requires fast shallow perceptual judgments). A relevance-based prioritization mechanism that coheres with the Relevance Principle should thus exclude this.

Of course, it's not that stereotype information could *never* be relevant to a judgment about a perceptual property. For example, we could imagine a task in which participants were told men faced left and women faced right and then had to guess peoples' facing direction—*without* seeing their pictures or knowing their genders—based solely on seeing their profession labels. In the absence of perceptual information, stereotype information could be relevant to participants' guesses as to what direction headshot subjects might be facing (e.g., learning men face left and seeing the "doctor" label might cause you to guess that the person would be facing left because you assume if he's a doctor he must be a man).

However, in this imagined experiment participants are making a different sort of orientation judgment than participants in our actual experiments made (recall that our participants judged the facing direction of headshot subjects *while looking at their headshots*). In the imagined experiment, participants are making judgments *about* a perceptual property (i.e., guessing what direction someone would be facing) in the absence of any relevant perceptual information (i.e., without seeing them) based entirely on non-perceptual information (e.g., seeing the "doctor" label and being told that male people face left). Therefore, in the imagined experiment, participants wouldn't be making *perceptual* judgments, they would be making judgments about non-present perceptual properties. Thus, we could gerrymander experiments wherein diverse kinds of information (including social information) could be relevant to non-perceptual judgments about perceptual properties. However, in our experiments, participants' orientation judgments are made on the basis of processing *present* perceptual information,

which means the judgments they made were perceptual. And the only information relevant to a perceptual judgment made in the presence of corresponding perceptual information is the information outputted by perceptual processing.

Nonetheless, our findings suggest that irrelevant stereotype information is being prioritized at the expense of perceptual information. This demonstrates the surprising power of prioritized stereotypes. Seeing a person and judging that they are facing left should be a straightforward shallow perceptual judgment, immune from the influence of high-level social and conceptual categories. But we demonstrate that stereotype information can seep into the bedrock of cognition, infecting even our most basic and automatic judgments. In that sense our findings make the strongest case for the cognitive priority of stereotypes: we find stereotype prioritization at the *lowest* levels of cognition where we should *least* expect it, suggesting that the influence of stereotypes could be felt anywhere.[11]

4. Stereotypes and Processing Fluency

I've argued that, in violation of the Relevance Principle, irrelevant stereotype information can become prioritized at the expense of task-relevant information even in our shallowest perceptual judgments. As I've hinted, prioritizing stereotype information has practical implications for discriminatory social practices. I want to conclude by gesturing at one particularly worrisome implication, which relates to *processing fluency*.

Processing fluency involves the subjective experience that accompanies relative ease of cognitive processing (Alter & Oppenheimer 2006). For example, in one of the first fluency studies Schwatz et al. (1991) asked participants to recall either 6 or 12 examples of their own assertive behavior, after which they rated their own assertiveness. People who only had to recall 6 examples rated themselves as more assertive than those who were asked to recall 12. The researchers argued that participants were using their experience of fluency as evidence of how assertive they actually were. Because the task of recalling 12 examples of assertive behavior was more difficult (and thus more disfluent) than recalling 6, those that had to recall 12 examples took their experience of disfluency as evidence that they weren't that assertive after all.

It has been since demonstrated that we use processing fluency as a cognitive heuristic for judging—among other things—truth (Reber & Schwarz 1999), preferability (Isengard & Lepper 2000; Oppenheimer 2006), and fittingness (Schwartz, Bless, Strack et al. 1991). That is, we take the phenomenology associated with fluent processing to be evidence of truth, preferability, and fittingness and the phenomenology associated with disfluent processing to be evidence of the opposite.

[11] Further, this violation of the Relevance Principle can't be explained in terms of computational limitations like the cognitive heuristic cases. Consider the availability heuristic again. We judge the probabilities of uncertain events according to the availability heuristic because events we can most easily bring to mind are *usually* the most relevant for accurately estimating probability. But it's simply not the true that stereotype information is usually more relevant to shallow perceptual judgments than perceptual information.

The reaction time difference we found between stereotype-congruent and incongruent trials suggests judgments about stereotype-congruent people are processed more fluently than judgments about stereotype-incongruent people. This difference in processing fluency is significant enough that many people consciously register it. For example, totally blind to the aim of the experiment, a number of participants in a debriefing afterwards spontaneously said they noticed a phenomenological difference between stereotype-congruent and stereotype-incongruent trials. To quote one participant: "the people that look like they would be doctors are not always, and the ones you definitely think are nurses are actually doctors" (Baker, Morales, & Firestone 2019). Having to make arbitrary perceptual judgments about stereotype-incongruent people *feels* more difficult, which suggests those judgments are being processed less fluently. You can experience this for yourself by taking one of our experiments online: https://perceptionresearch.org/stereotypes.

But should we be concerned that it takes us a little longer to judge something as irrelevant as what direction a person is facing? The literature on processing fluency gives us reason to think that we should. Again, when judgments about a person are experienced as comparatively disfluent, that disfluency can get interpreted as evidence that the person is less truthful, fitting, and preferable, even if the disrupted judgment (e.g., whether someone is facing left or right; the color of their shirt) is irrelevant to truth, fittingness, and preferability. For example, Reber and Schwarz (1999) found that sentences in less readable fonts were judged less true than the same sentences in more readable fonts. The cognitive effort we expend trying to decipher less readable fonts creates relative processing disfluency, which gets interpreted according to the fluency heuristic as the sentences being less truthful.[12] This sort of pattern suggests that experiencing processing disfluency when judging the facing direction of a stereotype-incongruent person (compared to a stereotype-congruent person) might subtly motivate us to regard them as less truthful and preferable. Thus, while it might seem unimportant that stereotypes make shallow perceptual judgments about stereotype-incongruent people comparatively disfluent, disfluency accompanying *any* judgment can trigger morally laden negative assessment.

What does all this mean for the Relevance Principle? It's important to tease apart two different things that are happening in our experiments. First, irrelevant stereotype information is being prioritized over relevant perceptual information, disrupting shallow perceptual judgments about stereotype-incongruent people. This is a straightforward violation of the Relevance Principle for reasons I laid out in the previous section. Second, when judgements about stereotype-incongruent people are disrupted, those judgments are experienced as more disfluent than judgements about stereotype-congruent people. This comparative disfluency can then subtly motivate negative evaluation of stereotype-incongruent people. But note that the fluency heuristic doesn't obviously violate the Relevance Principle (recall from section two that the Relevance Principle can accommodate for

[12] Which version would you trust more?

a. There are 8.4 million in New York City.
b. *There are 8.4 million people in New York City.*

cognitive heuristics).[13] However, negative evaluation of stereotype-incongruent people is a direct result of stereotype prioritization violating the Relevance Principle and thus seems worthy of mention.

Where does this leave us? Stereotype prioritization makes judgments about stereotype-incongruent people disfluent, which has implications beyond the particular disrupted judgments. In fact, the principal harm of stereotype prioritization often isn't the disruption of the judgment itself—e.g., taking longer to judge what direction a stereotype-incongruent person is facing. Rather, the harm comes from the accompanying feelings of disfluency which motivate unwarranted negative assessment of the stereotype-incongruent person's character—e.g., that they aren't honest. For example, experiencing disfluency when judging a female doctor's facing direction or the color of her shirt might make her seem less fitting in her social role. We often experience this type of disfluency as a feeling of general discomfort, which can motivate us to avoid interactions with stereotype-incongruent people altogether (e.g., "the appointment felt a little *off*, I'd just feel more comfortable going to this other [male] doctor").

Further, we tend to wrongly assume that the discomfort we experience interacting with stereotype-incongruent people isn't caused by stereotyping, especially if the judgments experienced as disfluent aren't relevant to the content of any stereotype (e.g., what direction someone is facing). Rather, the discomfort gets interpreted as something *specific* to the stereotype-incongruent person that is incidental to their stereotype-incongruent identity (e.g., "we're not a good personal fit"; "she just doesn't seem as knowledgeable", etc.). Therefore, while we consciously register that individual stereotype-incongruent people make us feel less comfortable and seem less trustworthy than individual stereotype-congruent people, we often fail to consider that these "individual" evaluations might in fact stem from systemic prejudices.

Lastly, stereotype prioritization coupled with the fluency heuristic can help explain why many people still prefer to see male doctors (or white professors, straight politicians, etc.) while espousing egalitarian beliefs. When stereotype information is prioritized, even arbitrary judgments about stereotype-congruent male doctors (white professors, straight politicians, etc.) are experienced as more fluent, which we erroneously interpret as them in fact being more competent than their stereotype-incongruent counterparts. Therefore, prioritizing stereotype information subtly motivates us to engage in preferential action towards stereotype-congruent people and discriminatory actions towards stereotype-incongruent people because engagement with stereotype-congruent people is experienced as more fluent and comfortable.

[13] Because the fluency heuristic is a heuristic, we might assume that the Relevance Principle would be able to accommodate for it. Perhaps processing fluency is *usually* indicative of truthfulness and preferability, such that fluent information is usually *relevant* to judging truth and preferability (see Unkelbach 2006 for this sort of account). If this were true, then defenders of the Relevance Principle can argue that we reason according to the fluency heuristic because it usually prioritizes the most relevant information.

Conclusion

I have shown that stereotypes are cognitively privileged in that stereotype information appears to be systematically prioritized over more relevant information, contrary to the basic assumption that our cognitive architecture is organized around the prioritization of relevance. Specifically, our empirical findings demonstrate how prioritized stereotypes can hijack automatic cognitive mechanisms like statistical learning, disrupting even our most basic perceptual judgments at the shallowest levels of cognition. Finally, stereotype prioritization can make even seemingly benign interactions with stereotype-incongruent people feel disfluent, motivating us to unknowingly discriminate against them in virtue of their stereotype-incongruent identities.

Bibliography

Aiden, H. S., & McCarthy, A. (2014). Current attitudes towards disabled people. *Scope*. http://www.scope.org.uk/About-Us/Media/Press-releases/May-2014/New-research-Majority-ofBrits-uncomfortable-talki

American Academy of Pediatrics. *Periodic Survey of Fellows*. Available at: http://www.aap.org/research/periodicsurvey/. Accessed Sept 19, 2020.

Apps, M., McKay, R., Azevedo, R. T., Whitehouse, H., & Tsakiris, M. (2018). Not on my team: Medial prefrontal cortex responses to ingroup fusion and unfair monetary divisions. *Brain and Behavior, 8(8)*, e01030.

Attardo S. (2016). Context as relevance-driven abduction and charitable satisficing. *Frontiers in Psychology, 7*, 305.

Bauer, P. J. (1993). Memory for gender-consistent and gender- inconsistent event sequences by twenty-five-month-old children. *Child Development, 64*, 285–297.

Baker, A., Morales, J. & Firestone, C. (2019). You're my doctor: Stereotype-incongruent identities impair recognition of incidental visual features. *Journal of Vision, 19*(10), doi: 10.1167/19.10.156.

Belle, D., Tartarilla, A., Wapman, M., Schlieber, M. & Mercurio, A. (2021). I can't operate, that boy is my son!": Gender schemas and a classic riddle. *Sex Roles*.

Bendana, J. & Mandelbaum E. (forthcoming). The fragmentation of belief. In *The Fragmented Mind*, eds. Borgoni, C., Kindermann, D. & Onofri A. Oxford: Oxford University Press.

Bian, L., Leslie, S.-J., & Cimpian, A. (2018). Evidence of bias against girls and women in contexts that emphasize intellectual ability. *American Psychologist, 73(9),* 1139–1153.

Bonjour, L. (1985). *The Structure of Empirical Knowledge*. Cambridge, MA: Harvard University Press.

Brosch, T., & Sharma, D. (2005). The role of fear-relevant stimuli in visual search: A comparison of phylogenetic and ontogenetic stimuli. *Emotion, 5(3*), 360–364.

Bublatzky, F., Pittig A., Schupp, H. T., & Alpers, G. W. (2017). Face-to-face: Perceived personal relevance amplifies face processing, *Social Cognitive and Affective Neuroscience, 12(5*), 811–822.

Carrasco, M. (2011). Visual Attention: The past 25 years. *Vision Research, 51(13*), 1484–1525.

Constable, M., Elekes, F., Sebanz, N., & Knoblich, G. (2019). Relevant for us? We-prioritisation in cognitive processing. *Journal of Experimental Psychology: Human Perception and Performance, 45(12)*, 1549–1561.

Cox, W. T. L. & Devine, P. G. (2015). Stereotypes possess heterogeneous directionality: A theoretical and empirical exploration of stereotype structure and content. *PLoS One, 10(3*), e0122292.

Dennett, D. (1978), *Brainstorms*, Cambridge: MIT Press.

Dotsch, R., Hassin, R. & Todorov, A. (2017). Statistical learning shapes face evaluation. Nature Human Behavior 1, 0001.

Dovidio, J. F., Hewstone, M., Glick P., & Esses V. M. (2010). Stereotyping and discrimination: Theoretical and empirical overview. In *The SAGE Handbook of Prejudice, Stereotyping and Discrimination*, eds. J. Dovidio, M. Hewstone, P. Glick, & V. Esses. London: SAGE Publications Ltd., 3–28.

Early, A. & Chaiken, S. (1998). Attitude structure and function. In *The Handbook of Social Psychology*, eds. D. Gilbert, S. Fiske, & G. Lindzey. New York: Oxford University Press, 269–322.

Egan. A. (2008). Seeing and believing: Perception, belief formation and the divided mind. *Philosophical Studies, 140(1)*, 47-63.

Elga, A. & Rayo A. (forthcoming). Fragmentation and information access. In *The Fragmented Mind*, eds. Borgoni, C., Kindermann, D. & Onofri A. Oxford: Oxford University Press.

Fiske, S., & Taylor, S. (1984). *Social Cognition*. New York: McGraw Hill.

Fodor, J. A. (1983). *The Modularity of Mind*. Cambridge: MIT Press.

Fodor, J. A. (1987). Modules, frames, fridgeons, sleeping dogs, and the music of the spheres. In *The Robot's Dilemma: The Frame Problem in Artificial Intelligence*, ed. Pylyshyn, Z. W., New York: Ablex.

Glaeser, E. (2005). The political economy of hatred. Quarterly Journal of Economics, 120, 45–86.
Hamlin, J. K., Wynn, K., & Bloom, P. (2007). Social evaluation by preverbal infants. *Nature, 450(7169)*, 557-559.

Hilton, J. L. & von Hippel, W. (1996). Stereotypes. *The Annual Review of Psychology, 47*, 237–271.

Iyengar, S. & Lepper, M. (2000). When choice is demotivating: Can one desire too much of a good thing? *Journal of Personality and Social Psychology, 79(6)*, 995–1006.

Jackson, L. M. (2020). *The Psychology of Prejudice: From Attitudes to Social Action* (2nd ed.). American Psychological Association.

Judd, C. M., Park, B., Ryan, C. S., Brauer, M., & Kraus, S. (1995). Stereotypes and ethnocentrism: Diverging interethnic perceptions of African American and White American youth. *Journal of Personality and Social Psychology, 69*, 460–481.

Jussim, L., Crawford, J. T., & Rubinstein, R. S. (2015). Stereotype (In)Accuracy in Perceptions of Groups and Individuals. *Current Directions in Psychological Science, 24(6)*, 490–497.

Kahneman, D. (2003). A perspective on judgment and choice: Mapping bounded rationality. *American Psychologist, 58(9)*, 697–720.

Kanwisher, N., & Yovel, G. (2006). The fusiform face area: a cortical region specialized for the perception of faces. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences, 361(1476)*, 2109–2128.

Lewis, D. (1982). Logic for equivocators. *Noûs*, *16(3)*, 431–441.

Lyons, J. (2005). Perceptual belief and nonexperiential looks. *Philosophical Perspectives, 19*, 237–256.

Mandelbaum, E. (2018). Seeing and conceptualizing: Modularity and the shallow contents of perception. *Philosophy and Phenomenological Research, 97*, 267–283.

Mandelbaum, E. (2016). Attitude, Inference, Association: On the Propositional Structure of Implicit Bias. Noûs, *50*, 629–658.

Monroy C., Meyer M., Gerson S., Hunnius S. (2017). Statistical learning in social action contexts. *PLoS ONE 12(5)*, e0177261.

Oppenheimer, D. (2005). Consequences of erudite vernacular utilized irrespective of necessity: problems with using long words needlessly. *Applied Cognitive Psychology, 20(2)*, 139–156.

Petty, R. E. & Wegener, D. T. (1998). Attitude change: Multiple roles for persuasion variables. In *The Handbook of Social Psychology*, eds. D. Gilbert, S. Fiske, & G. Lindzey. New York: Oxford University Press, 323–390.

Potter M. C., Wyble B., Hagmann C. E., & McCourt E. S. (2014). Detecting meaning in RSVP at 13 ms per picture. *Attention Perception Psychophysics, 76(2*), 270–279.

Quilty-Dunn, J. (2019). Reid's doxastic theory of perception. *Philosophers Imprint, 19(31*), 1–11.

Quilty-Dunn, J., Mandelbaum, E. (2018). Against Dispositionalism: Belief in Cognitive Science. *Philosophical Studies 175*, 2353–2372.

Reber, R., & Schwarz, N. (1999). Effects of perceptual fluency on judgments of truth. *Consciousness and Cognition: An International Journal, 8(3*), 338–342.

Romberg, A. R., & Saffran, J. R. (2010). Statistical learning and language acquisition. *Wiley interdisciplinary Reviews: Cognitive Science, 1(6*), 906–914.

Ruble, D N.., Martin, C., & Berenbaum, S. (2006). Gender development. In *Handbook of Child Psychology: Vol. 3. Social, Emotional, and Personality Development*, ed. N. Eisenberg. Hoboken, NJ: Wiley, 858–932.

Saffran, J., Aslin, R., & Newport, E. (1996). Statistical learning by 8-month-old infants. *Science, 274(5294*), 1926–1928.

Saxe R., & Kanwisher N. (2003). People thinking about thinking people. The role of the temporoparietal junction in "theory of mind". *Neuroimage. 19(4*), 1835–1842.

Schneider, D. J. (2004). *The Psychology of Stereotyping*. New York: Guilford Press.

Schröder T., Rogers K. B., Ike S., Mell J. N., & Scholl W. (2013). Affective Meanings of Stereotyped Social Groups in Cross-Cultural Comparison. *Group Processes & Intergroup Relations. 16(6*), 717–733.

Schwarz, N., Bless, H., Strack, F., Klumpp, G., Rittenauer-Schatka, H., & Simons, A. (1991). Ease of retrieval as information: Another look at the availability heuristic. *Journal of Personality and Social Psychology, 61(2*), 195–202.

Seigel, S. (2017). *The Rationality of Perception*. Oxford: Oxford University Press.

Simon, H. (1957). Models of Man; Social and Rational. Oxford: Wiley.

Smiley, C., & Fakunle, D. (2016). From "brute" to "thug:" the demonization and criminalization of unarmed Black male victims in America. *Journal Of Human Behavior In The Social Environment, 26(3-4)*, 350–366.

Spelke E. S., & Kinzler K. D. (2007). Core knowledge. *Developmental Science,10(1*), 89–96.

Sperber, D., & Wilson, D. (1995). *Relevance: Communication and Cognition (2nd ed.)*. Oxford: Blackwell Publishing.

Sperber, R., McCauley, C., Ragain, R.D. et al. (1979). Semantic priming effects on picture and word processing. *Memory & Cognition, 7*, 339–345.

Stalnaker, R. (1984). *Inquiry*. Cambridge: MIT Press.

Steele, C. M. (2010). *Whistling Vivaldi: How Stereotypes Affect Us and What We Can Do*. New York: W. W. Norton & Co.

Trautner H. M., Gervai J., & Németh R. (2003). Appearance–reality distinction and development of gender constancy understanding in children. *International Journal of Behavioral Development, 27(3*), 275–283.

Tversky, A., & Kahneman, D. (1974). Judgment under Uncertainty: Heuristics and Biases. *Science, 185(4157*), 1124–1131.

Unkelbach C. (2006). The Learned Interpretation of Cognitive Fluency. *Psychological Science*, *17(4*), 339-345.

Valian, V. (1998). *Why so Slow? The Advancement of Women*. Cambridge: MIT Press.

Valian, V. (2005). Beyond gender schemas: Improving the advancement of women in academia. *Hypatia, 20*, 198–213.

Vickery, T. J., Park, S. H., Gupta, J. et al. (2018). Tasks determine what is learned in visual statistical learning. *Psychonomic Bulletin Review 25*, 1847–1854.

Wilbourn, M. P. & Kee, D. W. (2010). Henry the Nurse is a Doctor Too: Implicitly Examining Children's Gender Stereotypes for Male and Female Occupational Roles. *Sex Roles, 62*, 670–683.

Wilson J. W., Hugenberg K., & Rule, N. (2018). Racial bias in judgments of physical size and formidability: From size to threat. *Journal of Personality and Social Psychology, 113(1*), 59–80.

Wilson, D. & Sperber, D. (2004). Relevance Theory. In *The Handbook of Pragmatics*, eds. L. R. Horn & Ward, G. Oxford: Blackwell, 607–632.