“You’re my doctor?”
Stereotype-incongruent identities impair recognition of incidental visual features

Austin A. Baker, Jorge Morales, & Chaz Firestone

Abstract

Stereotypes shape our judgments about people around us—as when women are assumed to be students, research assistants, or nurses rather than professors, principal investigators, or doctors. Can stereotypes also intrude on representations that have nothing to do with the content of the stereotype? Here, we explore how the assumptions we make about other people impair our ability to process completely incidental, and surprisingly low-level, aspects of their appearance—including even their location in space. We showed subjects headshots of male and female medical professionals, and asked them simply to indicate the direction of the target's shoulders (left or right)—an extremely straightforward task that subjects performed with near-ceiling accuracy. The key manipulation was a cue on each trial that the upcoming image would be of a "doctor" or a "nurse", and a statistical regularity in the experiment such that "doctor"-labeled images tended to face one way and "nurse"-labeled images tended to face the other way. Although gender was completely irrelevant to any aspect of the task, subjects were slower to judge the orientation of stereotype-incongruent people (female "doctors" and male "nurses") than stereotype-congruent people (male "doctors" and female "nurses"), even though the images were labeled arbitrarily. Follow-up experiments showed that this effect couldn't be explained by the raw surprisingness of, e.g., seeing a man when expecting a nurse; instead, these results suggest that even straightforward forms of statistical learning can be intruded upon by long-held social biases, in ways that alter processing of incidental, basic visual features.

Keywords: stereotypes, gender, perception, statistical learning, social bias, implicit attitudes

Consider this popular "riddle":

> *A father and his son are in a horrible car accident where the father is tragically killed. The boy is rushed to the hospital with critical injuries and is wheeled into the operating room, where a distinguished surgical team is assembled. However, upon seeing the boy, the lead surgeon exclaims, "I can't operate on him, he's my son!".*
>
> *How can this be?*

When this riddle is told to a group of people, many of them may struggle to answer it; perhaps the boy was adopted, or perhaps the surgeon is his stepfather (and the boy had been riding with his biological father). Of course, the "solution" is simpler: the surgeon is the boy's *mother*. The riddle thus highlights a stereotype about the relationship between gender and profession: namely, that a doctor tends to be, or even *should* be, a man, whereas other medical professions (such as nursing) are more "appropriate" for women (Cann, 1993; Levy, Taylor & Gelman, 1995; Wilbourn & Kee, 2010).

Stereotypes are mental schemas we have about groups of people and their qualities — "more than just beliefs about groups, they are also theories about how and why certain attributes go together" (Hilton & von Hippel, 1996). Stereotypes may thus carry diverse kinds of social information (including the social roles that group members are expected to occupy; Dovidio, Hewstone, & Glick, 2010; Lippman, 1922) and can lead to biases in the attitudes and actions we take towards other people (Eagly & Chaiken, 1998; Fiske, 1998). For example, stereotypes based on race, gender, nationality, sexuality, and many other such attributes can affect how competent we rate others as being (Bodenhausen, Kang, & Peery 2012; Fiske, Cuddy, Glick, & Yu, 2002), how much we are willing to pay them as employees (or whether we will hire them in the first place; Moss-

Rascusin, Dovidio, Brescoll, Graham, & Hangelsman, 2012), whether we would vote for them in elections (Chiao, Bowman, & Gill, 2008), and even what internal mental experiences we imagine them to be having (as in, for example, stereotypes about how different racial groups experience pain; Trawalter, Hoffman, & Waytz, 2012) — among many more such effects. Moreover, individuals whose identities defy the stereotypes we hold may be judged negatively as a result; for example, a male nurse, female doctor, or other individual whose gender, race, nationality, or sexuality seems to "misalign" with some other social role they hold may be judged suspiciously or negatively compared to someone whose identities conform to our stereotypes (Stern, West, & Rule, 2015).

What is the "Reach" of Stereotypes?

On one hand, the scope of influence of stereotypes is impressively (and depressingly) broad, as shown by the very diverse consequences that stereotypes may have. On the other hand, however, such consequences are also *narrow*, in at least one key sense: they all concern judgments, behaviors, and responses that fall *within the domain of the stereotype's content*. For example, when employers choose not to hire applicants from a given racial group, their biased behaviors arise from a contentful *connection* (whether implicit or explicit) between the judgment they are making — here, to hire or not to hire — and the stereotype they hold (e.g., that members of certain groups are less productive, or even less intelligent). Similarly, when voters choose not to elect candidates of a certain gender, their biased preference is explained in terms of biased beliefs about whether people of a certain gender are fit for office (perhaps because of stereotypical beliefs about the emotional states of men vs. women). In other words, in these cases the *domain* of the stereotype is the same as the domain of the biased judgments that result from it, in that stereotypical beliefs about groups of people and attributes affect subsequent judgments about those people and attributes.

Indeed, even research on more implicit or automatic stereotype biases still tends to explore judgments that fall within the relevant stereotype’s domain. For example, biased responding on the Implicit Association Test (Greenwald, McGhee, & Schwartz, 1998) involves automatic and implicit processing that differs from the sorts of explicit judgmental biases reviewed above (cf. Mandelbaum, 2016; Melnikoff & Bargh, 2018). But even implicit biases of this sort operate through a conceptual connection between the stereotype the subject holds and the judgment they make. For example, if implicit stereotypes cause American subjects to be faster at grouping photographs of non-white targets with images of “foreign” landmarks (Nosek, Smyth, Hansen, Devos, Lindner et al., 2007), then the stereotypes (e.g., representing non-white people as somehow un-American) are still affecting judgments that fall within the sphere of their conceptual domain. Similarly, some research shows that priming subjects with one racial identity or another (e.g., African-American vs. Caucasian) can cause subjects to misidentify a tool as a weapon (Payne, 2006). But this response too arises because of the connection between the stereotype and the response — e.g., because of stereotypes connecting race to violent behaviors — and so again is an example of a judgment that falls *within* the stereotype’s domain.

Of course, it is completely natural, unproblematic, and unsurprising that most of the work on stereotypes to date has focused on biases that fall within the stereotypes’ content domains. However, what remains unclear from previous research is whether the “reach” of stereotypes must be constrained in this way, or instead whether the causal influence of stereotypes might extend *beyond* their domain. In other words, might stereotypes have a wider reach than is traditionally assumed?

For example, suppose you encounter someone with a “stereotype-incongruent” identity, such as a male nurse or a female doctor. In addition to judging them as less suited for their profession (which would be a “within-domain” type of bias), or judging them negatively in a general way (Stern et al., 2015), could other aspects of your representation of them be affected by this stereotype — including aspects of them that are *completely unconnected* to the stereotype's domain? For example, might you find it more difficult to see the color of their shirt, or the shape of their glasses? Here, we explore this sort of possibility. We ask whether stereotypes can affect our behavior in domains that go beyond the content of the stereotype in question, and even whether this influence can reach down into more basic cognitive mechanisms, including mechanisms of visual attention and perceptual learning.

Statistical Learning and Stereotyping

What kind of process might give rise to such "wide" reach? One route to such effects — the one we explore here — is *statistical learning* of visual features. Statistical learning is a pervasive and general cognitive process by which we extract regularities from our environment, often in the service of forming representations that reflect those regularities. A core domain of statistical learning research is the processing of speech sounds by infants and young children. For example, when exposed to sequences of syllables whose transitions obey certain statistical regularities (e.g., *BA-DI-KU-PA-BI-DU-LA-NI-PU-BA-DI-KU-LA-NI-PU-PA-BI-DU*), infants become sensitive to these transition probabilities, and they come to represent this structure (e.g., that the sequence is actually composed of the "words" *BADIKU*, *LANIPU*, and *PABIDU*; Saffran, 2003; Saffran, Aslin, & Newport, 1996).

Importantly, statistical learning also occurs in the visual domain, and in adults as well as infants. For example, if subjects are exposed not to sequences of sounds but instead to visual symbols (e.g., ♻-◎-∞-Δ-❄-∇-∂-◐-♦-♻-◎-∞-∂-◐-♦-Δ-❄-∇), they will come to represent the statistically co-occurring sub-sequences as distinct units (e.g., ♻◎∞, Δ❄∇, and ∂◐♦). Statistical learning may even occur over visual features that carry more social significance, such as facial expressions and other social cues (Phillips, Weisbuch, & Ambady, 2014; Ruffman, Taumoepeau, & Perkins, 2012).

Here, we suggest that statistical learning of this sort can *extend* the reach of stereotypes in a previously unexplored way. In particular, we ask whether this mechanism could allow stereotypes to shape surprisingly low-level representations that fall completely outside the content domain of the stereotype — including extremely basic and entirely arbitrary perceptual attributes, such as a person's physical location in space.

The Present Experiments: Statistical Learning over Stereotype-Incongruent Identities

The present experiments introduce a new paradigm to explore the potential for stereotypes to exert wider reach than their typical content domain. We investigated whether stereotypes connecting a person's gender to their profession (e.g., doctor–man and nurse–woman) could impair low-level perceptual judgments about the spatial orientation of the person's body. We collected headshots of male-presenting and female-presenting medical professionals (which we randomly labeled "doctors" and "nurses") and asked subjects to indicate which direction the target's shoulders were facing. Crucially, we introduced a simple and reliable statistical regularity to the stream of images: We manipulated the headshots' orientations so that all "doctors" (half of whom were men and half of whom were women) faced one way (e.g., left), and all "nurses" (half men and half women) faced the other way (e.g., right) — a regularity which could be exploited by subjects to facilitate quick and accurate responses. (For example, if you knew a "doctor" was coming up next, you could know which way they were going to face because of this regularity — whereas gender was completely uncorrelated with facing direction.) Would subjects nevertheless import gender stereotypes about what doctors and nurses should look like, in ways that impair their ability to incorporate these regularities? We hypothesized that they would — that subjects would be less able to capitalize on statistical regularities holding over female doctors and male nurses ("stereotype-incongruent" identities) than male doctors and female nurses ("stereotype-congruent" identities), as if the actual statistical regularities holding between profession and facing direction were really holding between *gender* and facing direction. As a result, we predicted slower reaction times when making orientation judgments about stereotype-incongruent individuals than stereotype-congruent individuals, but only when a statistical regularity was present. In other words, our studies ask whether stereotypes can intrude on these very basic mechanisms of perception, attention, and learning, in ways that allow their influence to reach far outside their conceptual domain.

Experiment 1: Stereotypes Intrude on Perceptual Learning and Recognition

Do stereotypes intrude on statistical learning of regularities in our environment, in ways that disrupt our ability to acquire and use such regularities for basic perceptual tasks? Experiment 1 tested this question using stereotypes between gender and profession as a case study.

Method

Open Science Practices

For this experiment and all others reported here, the raw data, analyses, materials, and experiment code are available at https://osf.io/6nhrd/.

Participants

One hundred subjects were recruited online via Amazon Mechanical Turk (for validation of this subject pool's reliability, see Crump, McDonnell, & Gureckis, 2013). Participants provided informed consent prior to participation and were financially compensated for their contribution.

Given the lack of previous research using this paradigm (which we introduce for the first time here), we chose a sample size of 100 subjects simply because it seemed *large*. However, note that Experiment 2 is a direct replication of the present experiment with a larger sample size determined based on a sensitivity power analysis using the results from Experiment 1, and that Experiment 3 also uses that same sample size, always with the same exclusion criteria. In this study, we report all measures, manipulations, and exclusions.

Materials

We obtained 60 standardized headshots (80x100px) of medical professionals (30 women, 30 men) from a major medical institution, which were validated in a separate study.[1] All the people depicted in the images had a salient shoulder facing direction (left or right), which we could manipulate in advance by reflecting the image about its vertical axis. (Due to the nature of online experiments, we can't be sure of the exact size, brightness, color [etc.] of these images as they actually appeared to subjects; however, any variation in such properties would be equated across our experimental conditions.)

Procedure

The experiment consisted of 60 trials, each of which proceeded as follows. First, the question "What's the direction of the DOCTOR's [NURSE's] shoulders?" was displayed above an empty frame; after 3 seconds, the target image then appeared inside that frame, with the question text still visible (Fig. 1). Subjects then had up to 2 seconds to indicate via keypress whether the target's shoulders faced left or right ("1"=left, "2"=right). Subjects received feedback on the correctness of their orientation response: the frame of the image turned green for correct responses and red for incorrect responses, for 500ms.

To ensure that subjects had encoded the stated profession of the medical professional, the trial continued by presenting subjects a new question (with the previous question now absent from

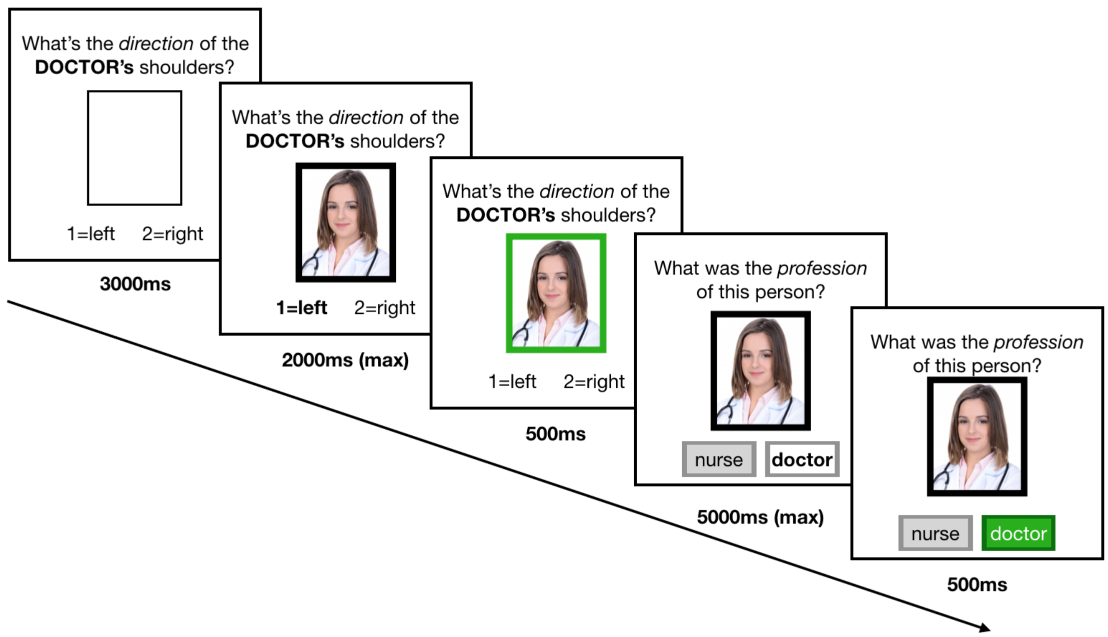
[1] We initially collected 82 headshots, and recruited 50 online subjects to judge whether the people in the images were facing left or right. Only the 60 images with the highest proportion of agreement about shoulder direction were included in the experimental stimuli (average agreement: 84%).

the screen): "What was the profession of this person?". Subjects now had 5 seconds to respond by clicking either a "nurse" or "doctor" button which appeared at the bottom of the screen (the order [left/right] in which the buttons appeared was randomized in each trial). Once they selected a profession, the button turned green (correct answers) or red (incorrect answers) for 500ms, and the trial ended. If subjects failed to provide either response within the time allotted, a message indicating that they were too slow appeared on the screen and they were asked to press the spacebar to continue. The next trial began 500ms after the conclusion of the previous trial. At the beginning of the experiment there were 6 practice trials to familiarize subjects with the task.

For each subject, half of the target images were arbitrarily labeled as "doctors" and the other half were labeled as "nurses" (which image received which labels was counterbalanced across subjects). Crucially, in Experiments 1 and 2 we manipulated the headshots' orientations so that 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 (counterbalanced across subjects). In other words, stated profession was predictive of facing direction but gender was not, which meant that subjects could capitalize on this regularity to facilitate their responses; once they knew whether a doctor or nurse was about to appear, they could also know which direction that person would be facing, if they had learned the statistical regularity. Readers can experience the task for themselves at https://osf.io/6nhrd/.

Figure 1

Task design for Experiments 1–3



Note. On each trial, subjects were first cued about the "profession" of an upcoming target ("doctor" or "nurse"); then, they judged the orientation of that person's shoulders and were asked to recall the person's labeled profession. In fact, all of the "doctors" (half men, half women) faced one way (e.g., left) and all of the "nurses" (half men, half women) faced the other way (e.g., right). We predicted that subjects would better utilize this regularity for male doctors and female nurses than for female doctors and male nurses, and so would be slower to report a female doctor or male nurse's orientation than a male doctor or female nurse's orientation. Readers can experience this task for themselves at https://osf.io/6nhrd/.

Exclusion Criteria

Given the simplicity of both tasks (i.e., determining the targets' shoulder direction and recalling the targets' profession), we expected attentive subjects who followed the instructions to perform with high accuracy. Thus, we determined a priori that only the data of subjects who completed both tasks with an accuracy of 80% or higher would be analyzed. Moreover, for computing subjects' reaction time averages in the main task (i.e., classifying the target's shoulder direction), we only included trials where subjects responded correctly to both questions. This was done to ensure that we only computed the reaction times in trials where we could confirm subjects paid attention to the professional label assigned to the target image, as this was the key manipulation in our experiments. Finally, trials faster than 100ms were excluded to avoid counting impossibly fast "lucky" correct responses. We applied these same exclusion criteria across all of the experiments reported here.

Results and Discussion

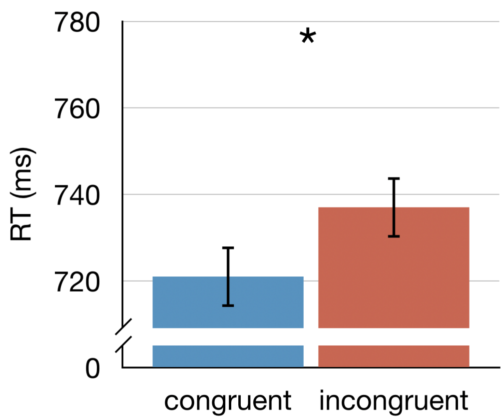
Subjects completed the task quickly and with a high degree of accuracy: Considering those subjects who passed our exclusion criteria, the average response time for classifying shoulder direction was 729ms, and the average accuracy was 97.2%, suggesting that subjects found the task relatively easy and straightforward (as expected). Subjects recalled the target's profession with an average accuracy of 96%.

Remarkably, however, stereotype-congruent trials showed a different pattern of responses than stereotype-incongruent trials. In particular, subjects were slower to judge the orientation of stereotype-incongruent headshots (female "doctors" and male "nurses") than stereotype-congruent headshots (male "doctors" and female "nurses"); 737ms vs. 721ms, $t(72)=2.31$, $p=.024$. In other words, even though the facing direction of the headshot was predicted only by the labeled profession and never by gender, subjects were slower to apply the profession regularity (e.g., doctors=left, nurses=right [randomly assigned for each subject]) for female "doctors" and male "nurses".

This suggests that the targets' gender impacted subjects' ability to exploit statistical regularities connecting profession to facing direction. Put differently, when target gender and profession cohered with stereotypical expectations (here, male doctors and female nurses), the connection between profession and facing direction better facilitated subjects' orientation judgments than when gender and profession did not cohere with stereotypical expectations (here, female doctors and male nurses) — even though gender had no statistical connection at all to facing direction. Thus, stereotypes connecting gender to profession intruded upon what might otherwise have been a straightforward process of learning and applying statistical regularities in one's environment.

Figure 2

Results from Experiment 1



Note. Average reaction times showing that subjects were faster to identify the target’s orientation in stereotype-congruent trials (female “nurses” and male “doctors”) compared to stereotype-incongruent (female “doctors” and male “nurses”) trials. Error bars indicate the s.e.m. of the difference between the two types of trials. * $p < 0.05$

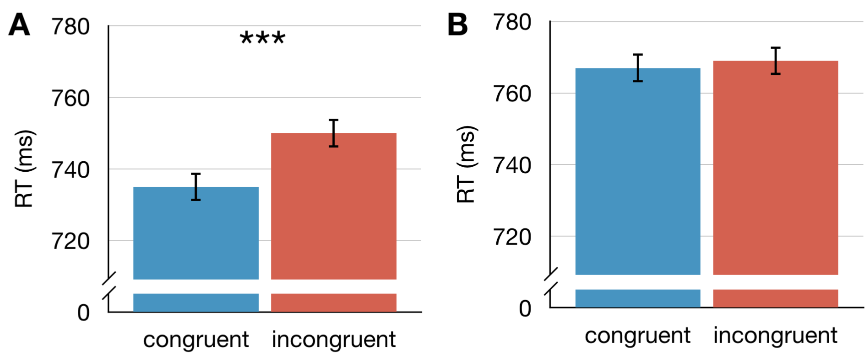
Experiment 2: Direct, High-Powered Replication

How robust is the finding from Experiment 1? Though that experiment showed a significant result, it was reliable at only p=.024. To be sure we can confidently rely on it (and to safely interpret our upcoming Experiment 3, which attempts to isolate the mechanism of this effect by trying to “eliminate” it), Experiment 2 directly replicated Experiment 1 with a larger sample: 300 subjects, instead of 100. Assuming a similar exclusion rate as in Experiment 1, this sample ensured 98% power to detect an effect of the same size as in Experiment 1.

Experiment 2 thus proceeded in the exact same way as Experiment 1; the *only* difference was the sample size. And indeed, we observed the same pattern of results: Subjects were slower to judge the orientation of stereotype-incongruent headshots (female “doctors” and male “nurses”) than stereotype-congruent headshots (male “doctors” and female “nurses”), and this time the effect was very reliable; 750ms vs. 735ms, $t(201)=3.80$, $p=.0002$. This encouraged us that the effects explored here are real and robust; and, crucially, it allowed us to be confident that any of our future studies that *fail* to find this effect (see Experiment 3) can be safely interpreted as null results.

Figure 3

Results from Experiment 2 & 3



Note. Average reaction times in (A) Experiment 2 and (B) Experiment 3 for identifying the target's shoulder direction in stereotype-congruent (female "nurses" and male "doctors") and stereotype-incongruent (female "doctors" and male "nurses") trials. Error bars indicate the s.e.m. of the difference between the two types of trials. *** $p < 0.001$. Whereas Experiment 2 showed an effect of stereotype congruence on reaction time (replicating the results of Experiment 1), Experiment 3 did not, ruling out alternative explanations based on "surprise".

Experiment 3: Ruling out "Surprise"

We have suggested that impaired responding to stereotype-incongruent targets (relative to stereotype-congruent targets) results from subjects failing to incorporate or apply the statistical regularities holding over profession and facing direction. However, an alternative explanation of these results that doesn't appeal to statistical learning at all is simply that, on any given trial, seeing a stereotype-incongruent target (e.g., a male nurse) is just *surprising*, and that this surprise alone impairs performance on that trial — regardless of any statistical regularities appearing in the preceding trials. In other words, maybe seeing a man when subjects were expecting to see a nurse just gives subjects pause all on its own, slowing their responses for that reason.

To rule out this alternative, we repeated Experiment 2 with every methodological detail held constant *except* the statistical regularity between profession and facing direction. In other words, it was still the case that: half of the headshots were labeled “doctor”, and half were labeled “nurse”; half of the doctors were men, and half were women; and half of the headshots faced left, and half faced right. However, the left or right facing directions were no longer correlated with profession; doctors and nurses (and men and women) were now equally likely to face left as right. Under this design, there was simply no statistical regularity to pick up on in the first place; however, it still *was* the case that half the trials were “stereotype-incongruent”, such that subjects would, for example, expect to see a nurse and then see a man. If brute “surprise” was driving the effects from Experiments 1–2, then we should observe the same reaction-time difference between stereotype-congruent and stereotype-incongruent trials here in Experiment 3. However, if, as we have suggested, these effects derive from intrusion of stereotypes on statistical learning *per se*, then the effects should disappear.

Results and Discussion

With no statistical regularity to be exploited in the first place, there was no difference in reaction time between stereotype-incongruent and stereotype-congruent trials; 769ms vs. 767ms, $t(211)=.30$, $p=.76$. In other words, subjects were no faster or slower to report the facing direction of male doctors and female nurses than they were to report female doctors and male nurses.

Importantly, given the highly reliable results from Experiment 2 and the identical sample size here in Experiment 3, we again had 98% power to discover an effect of this size. That we did not observe this effect suggests that this one change in design truly did eliminate the relevant effects. Indeed, not only was this effect statistically absent, it was also reliably smaller than the effect from Experiment 2 (two-sample t-test: $t(412)=2.44$, $p=.02$).

Another result from Experiment 3 that is consistent with our biased-statistical-learning account is slower reaction times in Experiment 3 compared to Experiments 1–2. In a post-hoc test comparing stereotype-congruent trials from Experiments 1–2 to the very same trials in Experiment 3, subjects responded faster overall in Experiments 1–2 than in Experiment 3 (two-sample t-test: $t(485)=2.37$, $p=.018$). If subjects in Experiments 1–2 capitalized on the statistical regularity between profession and orientation to facilitate their responding, then their reaction times should be faster than those of the subjects in Experiment 3, where there was no regularity to exploit in the first place. Though we do not rely too heavily on this result here (it was only an exploratory analysis, and our experiments were not designed to test this question), this additional result was encouraging for our overall interpretation that the results from Experiments 1–2 reflect biases in incorporating the statistical regularities in the experiments rather than the brute "surprise" of seeing a stereotype-incongruent person.

General Discussion

What is the "reach" of stereotypes? Both previous research and common sense seem to suggest that stereotypes influence only those judgments falling *within* the stereotype's domain: Intuitively, when we hold stereotypes about a group of people and their qualities or attributes, the kinds of judgments that will be biased are those connected to those attributes. However, here we suggest that stereotypes can reach *beyond* their content domain in a previously unexplored way, altering our judgments about completely incidental (and surprisingly low-level) aspects of someone's appearance.

<u>"Wide Reach" in the Lab and the World</u>

What does it mean — and why should it matter — that stereotypes exert biases "within" or "outside" their domains? We tend to think about stereotypes as being domain-restricted, affecting the judgments which are connected to their content. For example, we might presume that stereotyping doctors as male would cause us to negatively view the competence or reliability of female doctors. However, our work goes a step further and suggests that stereotypes can also exert surprisingly *wide* reach, impacting judgments that are conceptually far removed from their content domain. To our knowledge, this type of wide reach (e.g., stereotypes about the gender of doctors impacting the recognition of low-level and completely arbitrary perceptual features) is novel, and suggests that stereotypes may have a wider sphere of influence than has been previously demonstrated.

This wide reach may have practical consequences. One reason the study of stereotypes is not only interesting but also important is that such research may allow us to anticipate and identify the effects of bias in the real world. For example, knowing that race and gender can negatively impact professors' assessments of students' academic performance allows institutions to enact measures like anonymous grading to undermine the effects of gender and race biases (Brennan 2008). However, to counteract the effects of stereotyping, we first must know *where* such biases manifest in our cognitive lives — and in particular which of our judgments about people are susceptible to stereotypic influence. In demonstrating "wide" reach of stereotypes, our work suggests that stereotypes can affect even very arbitrary judgments — in other words, the kinds of judgments one might not already be on the lookout for. And whereas we study a seemingly innocuous and arbitrary consequence of such biases (since the effects we observe concern only the perceived orientation of a person, rather than their competence, warmth, or other socially important qualities), this arbitrariness is exactly why the relevant effects could be so pernicious; if even our ability to represent a property as arbitrary as someone's *facing direction* can be impaired by stereotypic biases, then just about *any* type of judgment should be vulnerable to this type of interference (especially since statistical learning is a pervasive, automatic, and domain-general process).

Stereotypes and Perception

Our work also adds to a growing literature exploring connections between stereotypes and more basic mechanisms of perception, attention, and recognition — though we make this connection in a different way than is typical. For example, previous work suggests that political or demographic beliefs about other people can change their perceived skin color, as when African-American faces appear darker than luminance-matched Caucasian faces (Levin & Banaji, 2006) or political conservatives judge Barack Obama's skin tone to be darker than political liberals do (Caruso, Mead, & Balcetis, 2009; see also Krosch & Amodio, 2014; Krosch, Bernsten, Amodio, Jost, & Van Bavel, 2013). However, it is not clear that such results have the consequences they may appear to have; for example, it is not clear whether they really do originate from stereotypes (Firestone & Scholl, 2015a) or whether they actually alter perception (Firestone & Scholl, 2015b). But even if such effects *do* genuinely reflect stereotypes biasing perception, all such effects still involve "within-domain" effects. For example, when beliefs about the skin tone of African-American faces cause those faces to appear darker, this occurs because of the straightforward connection between the stereotype and the conceptually relevant judgments.

By contrast, our results here differ from these past results in at least two ways. First, as noted above, our results explore effects of stereotypes *beyond* their hypothesized domain, rather than within that domain. But second, we propose here that an *indirect* mechanism (i.e., biased statistical learning) leads to such effects. Importantly, this mechanism — i.e., biased acquisition and use of information

— doesn't require a violation of the modularity or cognitively impenetrability of perception (Firestone & Scholl, 2016; Fodor, 1983; Pylyshyn, 1999). Unlike other hypothesized violations which require postulating changes to the contents of perception themselves, the hijacking of statistical learning processes by long-held stereotypes provides a parsimonious and straightforward explanation of our results that relies on well-understood mechanisms.

Consequences for Statistical Learning

These results may also have consequences for theories of statistical learning. Statistical learning is typically conceived and studied as an unbiased process that extracts regularities from the environment in a rote and straightforward manner (Saffran, 2003). By contrast, here we suggest not only that statistical learning of this sort might *underlie* various social biases (for discussion of that hypothesis, see Phillips et al., 2014), but also that the *evidence-gathering process itself* can be biased by prior (stereotypical) beliefs. In our studies, stereotypes about gender and profession intruded upon what might otherwise be a straightforward process of extracting statistical regularities, with more efficient use of those regularities for stereotype-congruent targets than stereotype-incongruent targets. Though this is consistent with how some researchers (including philosophers and social theorists; Siegel, 2013) account for the nature and origin of certain social biases, it is not typical for research on statistical learning to consider the prior biases that learners might have when extracting regularities from the environment. Future work may thus consider, or be sensitive to, such considerations in studying and theorizing about the nature of statistical learning itself.

In Conclusion

What is the reach of stereotypes? Using a novel paradigm, we found that stereotypes can piggyback off established mechanisms of attention and statistical learning to reach beyond their respective content domains, impairing judgments about completely irrelevant perceptual features. And if stereotypes can impact even such low-level and arbitrary perceptual judgments through domain-general statistical learning, then we might expect that almost *any* kind of judgment could be similarly vulnerable. These results thus suggest that stereotypes have the potential to shape — and corrupt — even our most basic (and most seemingly immune) judgments in profound and unexpected ways, suggesting a disconcertingly central role for bias in determining how we perceive and understand the world.

References

Bodenhausen, G. V., Kang, S. K., & Peery, D. (2012). Social Categorization and the Perception of Social Groups. In S. Fiske, & C. N. Macrae (Eds.), *The SAGE handbook of social cognition* (pp. 311–329). Los Angeles, CA: SAGE Publications Ltd. http://dx.doi.org/10.4135/9781446247631.n16

Cann, A. (1993). Evaluative expectations and the gender schema: Is failed inconsistency better? *Sex Roles, 28*(11-12), 667–678. http://doi.org/10.1007/BF00289986

Caruso, E., Mead, N., & Balcetis, E. (2009). Political partisanship influences perception of biracial candidates' skin tone. *Proceedings of the National Academy of Sciences*, *106*(48), 20168–20173. http://doi.org/10.1073/pnas.0905362106

Chiao, J. Y., Bowman, N. E., & Gill, H. (2008). The Political Gender Gap: Gender Bias in Facial Inferences that Predict Voting Behavior. *PLoS ONE, 3*(10), e3666–7. http://doi.org/10.1371/journal.pone.0003666

Crump, M. J. C., McDonnell, J. V., & Gureckis, T. M. (2013). Evaluating Amazon's Mechanical Turk as a Tool for Experimental Behavioral Research. *PLoS ONE*, *8*(3), e57410–18. http://doi.org/10.1371/journal.pone.0057410

Dovidio, J. F., Hewstone, M., Glick P., & Esses V. (2010). Stereotyping and discrimination: Theoretical and empirical overview. In J. F. Dovidio, M. Hewstone, P. Glick, & V. M. Esses (Eds.), *The SAGE handbook of prejudice, stereotyping and discrimination* (pp. 3–28). London: SAGE Publications Ltd. doi: 10.4135/9781446200919.n1

Eagly, A. H., & Chaiken, S. (1998). Attitude structure and function. In D. T. Gilbert, S. T. Fiske, & G. Lindzey (Eds.), *The handbook of social psychology* (pp. 269–322). New York, NY: McGraw-Hill.

Fiske, S. T. (1998). Stereotyping, prejudice, and discrimination. In D. T. Gilbert, S. T. Fiske, & G. Lindzey (Eds.), *The handbook of social psychology* (pp. 357–411). New York, NY: McGraw-Hill.

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. http://dx.doi.org/10.1037/0022-3514.82.6.878

Firestone, C., & Scholl, B. J. (2015a). Can you experience ‘top-down’ effects on perception?: The case of race categories and perceived lightness. *Psychonomic Bulletin & Review*, *22*(3), 694–700.

Firestone, C., & Scholl, B. J. (2015b). When do ratings implicate perception versus judgment? The “overgeneralization test” for top-down effects. *Visual Cognition*, *23*(9-10), 1217–1226.

Firestone, C., & Scholl, B. J. (2016). Cognition does not affect perception: Evaluating the evidence for 'top-down' effects. *Behavioral & Brain Sciences*, *e229*, 1–77. http://doi.org/10.1017/S0140525X15000965

Fodor, J. A. (1983). *Modularity of mind: An essay on faculty psychology*. Cambridge, MA: MIT Press. http://doi.org/10.1017/CBO9780511814273.046

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. http://doi.org/10.1037//0022-3514.74.6.1464

Hilton, J. L., & von Hippel, W. (1996). Stereotypes. *Annual Review of Psychology*, *47*, 237–271. http://dx.doi.org/10.1146/annurev.psych.47.1.237

Krosch, A. R., & Amodio, D. M. (2014). Economic scarcity alters the perception of race. *Proceedings of the National Academy of Sciences*, *111*(25), 9079–9084. http://doi.org/10.1073/pnas.1404448111

Krosch, A. R., Berntsen, L., Amodio, D., Jost, J., & Van Bavel, J. (2013). On the ideology of hypodescent: Political conservatism predicts categorization of racially ambiguous faces as Black. *Journal of Experimental Social Psychology*, *49*(6), 1196–1203. http://doi.org/10.1016/j.jesp.2013.05.009

Levin, D. T., & Banaji, M. R. (2006). Distortions in the perceived lightness of faces: The role of race categories. *Journal of Experimental Psychology: General, 135*(4), 501–512. http://dx.doi.org/10.1037/0096-3445.135.4.501

Levy, G. D., Taylor, M. G., & Gelman, S. A. (1995). Traditional and evaluative aspects of flexibility in gender roles, social conventions, moral rules, and physical laws. *Child Development*, *66*(2), 512–531.https://doi.org/10.1111/j.1467-8624.1995.tb00887.x

Lippmann, W. (1922). *Public Opinion*. New York, NY: MacMillan Co. http://dx.doi.org/10.1037/14847-006

Mandelbaum, E. (2016). Attitude, inference, association: On the propositional structure of implicit bias. *Nous*, *50*(3), 629–658. http://doi.org/10.1111/nous.12089

Melnikoff, D. E., & Bargh, J. A. (2018). The mythical number two. *Trends in Cognitive Sciences*, *22*(4), 280–293. http://doi.org/10.1016/j.tics.2018.02.001

Moss-Racusin, C. A., Dovidio, J. F., Brescoll, V. L., Graham, M. J., & Handelsman, J. (2012). Science faculty's subtle gender biases favor male students. *Proceedings of the National Academy of Sciences*,*109*(41), 16474–16479. http://doi.org/10.2307/41763373

Nosek, B. A., Smyth, F. L., Hansen, J. J., Devos, T., Lindner, N. M., Ranganath, K. A., … Banaji, M. R. (2007). Pervasiveness and correlates of implicit attitudes and stereotypes. *European Review of Social Psychology*. *18*(1), 36–88. http://doi.org/10.1080/10463280701489053

Payne, B. K. (2006). Weapon bias: Split-second decisions and unintended stereotyping. *Current Directions in Psychological Science*, *15*(6), 287–291. http://doi.org/10.2307/20183137

Phillips, L. T., Weisbuch, M., & Ambady, N. (2014). People perception: Social vision of groups and consequences for organizing and interacting. *Research in Organizational Behavior*, *34*, 101–127. http://doi.org/10.1016/j.riob.2014.10.001

Pylyshyn, Z. (1999). Is vision continuous with cognition? The case for cognitive impenetrability of visual perception. *Behavioral & Brain Sciences*, *22*(3), 341–365. http://doi.org/10.1017/S0140525X99002022

Ruffman, T., Taumoepeau, M., & Perkins, C. (2012). Statistical learning as a basis for social understanding children. *British Journal of Developmental Psychology*, *30*(1), 87–104.

Saffran, J. R. (2003). Statistical language learning: Mechanisms and constraints. *Current Directions in Psychological Science*, *12*(4), 110–114. http://doi.org/10.2307/20182854

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*,*274*(5294), 1926–1928. http://doi.org/10.1126/science.274.5294.1926

Siegel, S. (2013). The epistemic impact of the etiology of experience. *Philosophical Studies*, *162*(3), 697–722.

Stern, C., West, T., & Rule, N. (2015). Conservatives negatively evaluate counterstereotypical people to maintain a sense of certainty. *Proceedings of the National Academy of Sciences*, *112*(50), 15337–15342. http://doi.org/10.1073/pnas.1517662112

Trawalter, S., Hoffman, K. M., & Waytz, A. (2012). Racial bias in perceptions of others' pain. *PLoS ONE*, *7*(11), e48546–8. http://doi.org/10.1371/journal.pone.0048546

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: A Journal of Research*, *62*(9–10), 670–683. http://dx.doi.org/10.1007/s11199-010-9773-7