WHEN PERCEPTION BYPASSES TRUTH: ATTENTION, BIAS, AND THE STRUCTURE OF SOCIAL STEREOTYPES

by

ANNA AUSTIN BAKER

A dissertation submitted to the

School of Graduate Studies

Rutgers, The State University of New Jersey

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

Graduate Program in Philosophy

written under the direction of

Andrew Egan

and approved by

_____

_____

_____

_____

_____

New Brunswick, New Jersey

October 2019

ABSTRACT OF DISSERTATION

When Perception Bypasses Truth: Attention, Bias, and the Structure of Social Stereotypes

By ANNA AUSTIN BAKER

Dissertation Director:
Andrew Egan

Is perception accurate? How wide spread is inaccuracy in perception and under what conditions do our perceptual capacities undermine our ability to accurately perceive? This dissertation examines two examples of perceptual inaccuracy: attention altering perceptual phenomenology (making attended to stimuli appear bigger, brighter, and higher in spatial frequency) and social stereotypes impairing low-level perceptual judgments. There is a prevailing assumption in philosophy and cognitive science that perception is—and functions to be—truth oriented. However, I herein argue that our perceptual faculties often fail to deliver truth. Moreover, understanding *how* our cognitive architecture gives rise to systematic perceptual inaccuracy can provide us with insight into just how much our experience of the world is shaped by our social categories and computational limitations.

In chapters 1 and 2, I consider the way social stereotypes shape perceptual judgments. We know social stereotypes influence many of our judgments. Women, for example, are deemed less likely to succeed than men in especially intellectually demanding tasks (Bian et al. 2018). This suggests that higher-order judgments about qualities like 'brilliance' or 'genius' can be shaped by our gender stereotypes. But might stereotypes be so cognitively entrenched that they could affect more basic perceptual judgments as well? For example, would harboring the

stereotype 'doctors are men' make it more difficult to visually process a female doctor? These chapters empirically and philosophically consider this question and its larger social ramifications. I argue that my empirical work with Jorge Morales and Chaz Firestone suggests that stereotyping has a considerably wider scope of causal influence than has been appreciated in the philosophical and psychological literature, which can shed light of larger patterns of discrimination.

In chapter 3, I take on another, more basic, facet of perceptual inaccuracy—the phenomenological effects of voluntary and involuntary attention. I argue that much of the empirical evidence supports the interpretation that attention inaccurately distorts many aspects of our perceptual experience. On the face of it, these findings appear to be difficult to reconcile with the view that perception functions to furnish us with accurate representations of the world. However, rather than claim that our perceptual systems are constantly in the process of malfunctioning, I argue that perception instead functions to *guide action* and that this can satisfactorily explain many examples of perceptual inaccuracy.

ACKNOWLEDGEMENTS

I've been fortunate enough to have benefited from the friendship and advice of many people throughout the development and writing of this dissertation. First and foremost, all the members of my dissertation committee—Andy Egan, Susanna Schellenberg, Eric Mandelbaum, Chaz Firestone, and Alex Guerrero—whose guidance, kindness, and patience, over the years has immeasurably shaped me as a person and as a researcher. I also owe a great deal to the community of past and present graduate students at Rutgers (who managed to make philosophy and New Jersey fun!), in particular D Black, Laura Callahan, Megan Feeney, Will Fleisher, Carolina Flores, Nate Flores, Danny Foreman, Georgi Gardiner, Veronica Gomez, Jimmy Goodrich, E. J. Green, Chris Hauser, Caley Howland, Anton Johnson, Morgan Moyer, Dee Payton, Eli Shupe and Isaac Wilhelm. I am furthermore grateful for Brian McLaughlin, Karin Stromswold, and Sara Pixley at the Rutgers Center for Cognitive Science, who encouraged and facilitated what would become an all-consuming passion for cognitive science.

Outside of Rutgers, my work as considerably benefited from the feedback of Ned Block, Marissa Carrasco, and Jesse Prinz. I am also thankful for the comradeship of Robert Long, Jake Quilty-Dunn, Andrew Lee, Joseph Bendana, Zoey Lavelle, Caroline Bowman, Tyler Brook-Wilson, Daniel Young, and the rest of the New York philosophy community over the last five years.

Lastly, what was supposed to be a brief three-month stay in Baltimore ended up becoming one of the most intellectually and personally fulfilling years of my life, marked in particular by befriending and collaborating with Jorge Morales and Chaz Firestone. This project owes so much to faculty and graduate students at Johns Hopkins—especially Steven Gross, Justin Halberda, Johnathan Flombaum, Jason Fischer, Sarah Cormiea, Giulia Elli,

Chenxiao Guan, Alon Hafri, Judy Kim, Patrick Little, Cara Maritz, Jose Rivera-Aparicio, Zekun Sun, and Aditya Upadhyayula.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

Chapter 1:

The Reach of Bias:

Rethinking the Causal Profile of Social Stereotypes

1. Introduction

Most of us encountered some version of the following 'riddle' during our childhoods. 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 and prepared for an operation where a distinguished surgical team is assembled. 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 struggle with this question, arguing that perhaps the boy was adopted, that the surgeon is his stepfather, or that either the father or surgeon are imposters. Now consider a similar case. A mother and her daughter are in car accident and the mother is killed on impact. When her daughter is taken to the hospital the nurse exclaims 'I can't attend to her, she's my daughter!'. What's going on here? The surgeon is the boy's mother and the nurse is the girl's father. Astonishingly perhaps, Wapman and Belle (2014) found that only 15% of children (between the ages of 7 and 17) and 14% of Boston University students who had never heard the riddle before guessed that the right answers. Moreover, they found that factors which did not affect participants' ability to solve the riddle included: the participants' gender, their exposure to female physicians, their liberal/conservative political identification, and their score on the Modern Sexism Scale. Indeed, it's unsettlingly easy to feel the intuitive pull against the surgeon-as-mother/nurse-as-father responses; the word 'surgeon' has such strong male associations and the word 'nurse' has such strong female associations that even those of us who take ourselves to harbor

progressive views on gender roles experience frustration and embarrassment upon discovering that we were in the moment unable to imagine that the surgeon could be a woman and a nurse could be a man.

But we might also ask what other kinds of judgments can be shaped by our social stereotypes? To motivate this question, imagine yourself in Urgent Care. A woman walks up to you and introduces herself as 'Doctor Jane Smith'. How might your biases shape your perception of Doctor Smith? For one, it's easy to imagine that you might mistake her for a nurse. Perhaps she also might strike you as less professionally competent than her male counterparts, meaning that you would be more likely to second guess her diagnosis and solicit other opinions. But what about aspects of her person that are seemingly unrelated to the doctor/nurse gender stereotype (e.g. the color and texture of her hair, the design style of her office, the shape of her glasses, etc.)? In section 2, I will cash this question out in terms of *causal reach*: assuming the causal efficacy of social stereotypes how far can a stereotype's causal tentacles extend? My empirical work with Jorge Morales and Chaz Firestone in section 3 explores the reach question. I argue that our findings motivate a novel account of the structure and mechanism of social stereotyping. I conclude in section 4 by arguing that the picture of stereotypes laid out here is capable of shedding new light on larger patterns of discrimination.

2. What is a Stereotype?

Stereotyping is the cognitive component of the larger umbrella of 'social bias', which is typically broken down into the following three subcategories (Eagly & Chaiken 1998; Petty & Wegener 1998; Dovidio, Hewstone, Glick, & Esses 2010): (1) *prejudice* is an *affective* attitude (usually negative) towards a social group (e.g. not liking group X), (2) *stereotyping* is a set of *cognitive attitudes*, which associate a social group with negative characteristics (e.g. thinking X

people are lazy), and (3) *discrimination* is a pattern of motivated *behavior* toward a group (e.g. avoiding engagement with X people). Stereotypes can be helpfully understood as cognitive schemas, made up of an assortment of implicit and explicit[1] attitudes related to a social group and its members (see Hilton & von Hippel 1996 for more on schemas). Dovidio et al. (2010) emphasize the informational richness of stereotypes (7):

> Stereotypes not only reflect beliefs about the traits characterizing typical group members but also contain information about other qualities such as social roles, the degree to which members of the group share specific qualities (i.e. within-group homogeneity or variability), and influence emotional reactions to group members.

Stereotyping, therefore, involves bringing forth a set of social information, which I will refer to as the '*content* of the stereotype'. Of course, what sort of information gets included in the content will depend on the stereotype in question, as a sampling of recent empirical work reveals: women are judged to be less likely to succeed in intellectually demanding tasks than men (Bian, Leslie, & Cimpian 2018), young black men are perceived to be larger and more physically threatening than young white men (Wilson, Hugenberg, & Rule 2018), and disabled people are considered to be less productive in the workplace than nondisabled people (Aidan & McCarthy 2014). Furthermore, evidence suggests we start acquiring the contents of stereotypes early in childhood development. Children acquire gender role stereotypes by the age of 2 (e.g. 'boys grow up to be doctors', 'girls like pink', etc) (Bauer 1993; Miller, Trautner, & Ruble 2006; Wilbourn & Kee 2010) and between the ages of 2 and 6 children start to more rigidly apply patterns of stereotypic thinking in their normative prescriptions (e.g. 'he shouldn't

[1] As Pinal and Spaulding (forthcoming) note, *explicit* bias is pretty straightforward to test: "you can just ask people what they think about various social groups and try to control for social desirability censorship" (4). But, as the doctor riddle demonstrates, *implicit* bias is more complicated and can be held by people who explicitly affirm egalitarian principles (De Houwer, Teige-Mocigemba, Spruyt, & Moors 2009).

be a nurse', 'she can't like trucks', etc.) (Miller et al. 2003).

On the standard view, the contents of a stereotype are activated when an individual from the stereotyped category is encountered (Allport 1954; Fiske 1998; Schneider 2004; and Gilbert & Hixon 1991). Activation involves the following three-step process: (1) a person is recognized as a member of some particular social group or category, (2) information about, and traits associated with, that social group or category (i.e. the content of the stereotype) are activated, and (3) judgments and interactions with the person influenced by the stereotype's content (Fiske 1998; Moskowitz, Li, & Kirk 2004; Schneider 2004; and, for critical discussion, Müller & Rothermund 2014). Thus, upon encountering a person, a host of socially laden information is triggered, which then goes on to shape subsequent interactions. But how are interactions shaped?

Stereotypes simultaneously enhance and restrict. On one hand, the content of a stereotype contains a wealth of information that goes beyond what is directly observed, which can inform social interactions. But once a stereotype has been activated, "stereotype consistent characteristics are attended to most quickly" (Dovidio et al. 2010) and a filtering process can occur, whereby stereotype incongruent information goes unnoticed or is discarded. From an epistemic perspective this makes stereotypes especially resistant to challenge and revision. For example, harboring the stereotype that African-Americans are less intellectually competent than Caucasians, will cause one to notice the failures and disregard the successes of African-American colleagues (and vice versa with Caucasians). Of course, one might reasonably wonder why we have stereotypes at all if they are so problematic and epistemically limiting. Why would our cognitive architecture be organized in such a way?

Interacting with the world necessitates reliance on stored information. Even basic perceptual interpretation involves use of stored conceptual categories. As Jessie Munton

(forthcoming) argues, there exists no one-to-one function from proximal to dismal stimulus; the relationship is rather many-to-many—"the same object may produce different retinal stimulation on different occasions, and conversely different objects may give rise to the same retinal input". A pattern of coffee cup retinal stimulation is compatible with seeing a real coffee cup, seeing coffee cup façade, etc. To navigate this uncertainty the visual system utilizes stored background information to interpret incoming visual input and settle on a distal interpretation.[2] Use of background information also extends beyond vision. We also use background information to interpret complex social situations. If I spot a good friend from afar in a coffee shop getting a coffee to-go and they fail to acknowledge me, I will assume they didn't see me or were in a hurry rather than that they are for some unknown reason angry with me (even though both interpretations are compatible with the information I have). Thus, in light of our processing limitations and the onerous task of sorting through a noisy and ambiguous world, our cognitive systems at many levels make use of stored category information.

Using background information also tends to be very efficient. When I encounter a fire alarm I've never seen, I don't treat it like a totally novel object. My stored information about the category 'fire alarm' enables me to interact with new fire alarms efficiently while expending little cognitive effort. Recognizing and interacting with fire alarms in this way is efficient, assuming my beliefs about fire alarms are accurate and unbiased. But it's easy to see how using stored information could be inefficient if I had been fed false information about fire alarms; the walk to the office would be considerably more complicated if I thought fire alarms shot bullets at random intervals. We can think of the negative social stereotypes (about race, gender,

[2] This process often modeled in a Bayesian framework (see Orlandi 2014; Rescorla 2015; Scholl 2005; Feldman 2015; Mamassian, Landy, & Maloney 2001).

sexual orientation, etc.) as being like my 'firm alarms shoot bullets' belief; in encountering a stereotyped individual (a woman, person of color, etc.), false and normatively problematic information is automatically brought forth, which can hinder our ability to accurately interact with them. It might be true that we have to rely on stored information. But using stored categorical information to resolve ambiguity is only as efficient as the categories we have—and when it comes to the domain of person perception, many of our categories (i.e. our stereotypes) are harmful and inaccurate. We can thus think of stereotype categories as piggybacking off of more general (and often innocuous) cognitive capacities.[3]

Now with a more fleshed out picture of social stereotypes on the table, let's return to the question of causal reach. Stereotypes are casually efficacious and can shaped the judgments we make about other people. But which judgments specifically? To frame this question, I want to introduce a distinction between what I will call '*narrow*' and '*wide*' causal reach.[4] Again, consider the doctor/nurse gender stereotype. Recall that stereotype activation involves making information within the content of the stereotype available. The content of the doctor/nurse gender stereotype might include (implicit or explicit) beliefs about female doctors being unable to perform well under pressure (in, say, the ER) or male nurses being unable to effectively comfort and empathize with patients. In my terminology, if a stereotype is impacting judgments related to its content, then it is exerting narrow reach.

[3] Granted, you could think stereotypes are a special kind of category and are handled by a social cognition module. The point here is that you don't *have* to stipulate any special cognitive architecture to accommodate social stereotypes.

[4] For those averse to 'content' talk, another way to frame the distinction between narrow and wide reach is in terms of an infection analogy: if we think of stereotyping as an infection, we might ask what sorts of judgments are susceptible to the infection and what judgments are immune.

Narrow reach: A stereotype *S*, the content of which is an information set $\{i_1 \ldots i_n\}$, is narrowly reaching if *S* is impacting a judgment(s) directly related to $\{i_1 \ldots i_n\}$.

Thus, if 'female doctors don't perform well under pressure' is part of the content of the doctor/nurse gender stereotype, the stereotype would be narrowly reaching by motivating the judgment 'Sally, a female doctor, isn't performing well under pressure'. Narrow causal reach is depicted by the red arrows in figure 1 below.

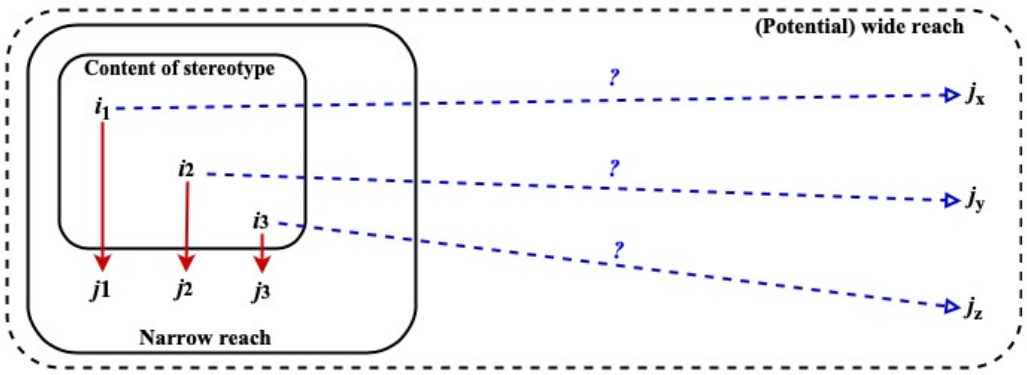

Figure 1. Narrow and wide causal influence. The stereotype's content includes the information set i_1, i_2, i_3 which are represented (with red arrows) as exerting *narrow* causal reach on judgments j_1, j_2, j_3, which are directly related to i_1, i_2, i_3 (e.g. if i_1 is 'doctors are men' then j_1 might be '*this* man is a doctor'[5]). Dotted blue arrows represent potential wide reach—the stereotype is (via some mechanism) causally influencing judgments (j_x, j_y, j_z) not directly related to its content.

However, what about judgments that fall outside of the stereotype's content? In the female doctor example this would be judgments unrelated to the gender stereotype—the shape of the doctor's glasses, the color of her hair, the architectural layout of her office, etc. In my

[5] The above table is a simplification of the causal structure. Of course, i_1 can influence more judgments that just j_1. If i1 is 'doctors are men' then it could influence any number of judgments: 'this man is a doctor', 'this woman is a nurse', etc.

terminology, if a stereotype is impacting judgments *not* related to its content, then it is exerting *wide* reach.

> *Wide reach*: A stereotype *S*, the content of which is an information set $\{i_1 \ldots i_n\}$, is widely reaching if *S* is impacting a judgment(s) *not* directly related to $\{i_1 \ldots i_n\}$.

But why should we think a stereotype would *ever* have wide reach? If wide reach was even possible what would the mechanism be for such causal influence? And would wide reach even matter in a larger sense (i.e. what would proving the existence of wide causal reach tell us about the nature of bias and the patterns of discrimination again stereotyped groups)? This paper aims to provide preliminary answers to these questions. I will argue that wide causal reach exists, it's potentially endemic, and it is able to elucidate important facets of social bias.

Before moving onto the empirical data, there are a couple things I want to flag about how narrow and wide reach have been defined here. First, I have not given an account of what exactly it means for a judgment to be "directly related" to a stereotype's content. I am intending to leave that question open to some degree of reasonable interpretation. Some judgments (for example, 'Sally, a female doctor, isn't performing well under pressure') are obviously related to the content of certain stereotypes and some judgments (for example, 'Sally, a female doctor, is wearing grey tennis shoes') are not. When describing narrow and wide reach I have tried to use what I take to be non-ambiguous examples. At the end of section 3 I will discuss more intermediate far reaching judgments, which are far reaching in that they aren't directly related to the content of the stereotype but potentially more closely connected to the content than some of the more extreme cases of far reach. Reach as I've specified it here is best understood as a *comparative* notion—some judgments will be nearer to a

stereotype's content than others. But we can set aside some of these in-between reach cases for the time being.

Second, the empirical work on stereotypes has almost exclusively focused on demonstrating individual instances of narrow reach. For example, given the content of race and gender stereotypes, the Wilson et al. (2018) finding that black men are judged to be more aggressive and the Bian et al. (2018) finding that women are judged to be less intellectually capable are clear examples of narrow reach. But there has been very little discussion on the cognitive architecture that supports stereotypes at all. Cox & Devine (2015) speak to this perceived deficit:

> Clear scientific progress toward understanding stereotyping and prejudice, therefore, requires a clear understanding of the cognitive architecture that underlies stereotypes. As Hilton and von Hippel lamented, however, many researchers' models and definitions of stereotypes and stereotyping are imprecise—or worse, unspecified—resulting in considerable ambiguity about the nature of stereotypes and stereotyping.

While stereotypes have been modeled in various ways (as prototypes, exemplars, schemas, etc.) little has been said about the underlying causal mechanisms of social stereotypes—particularly how they influence judgments and downstream patterns of discriminatory behavior (Cox & Devine 2015; Hilton & von Hippel 1996). This paper is meant, therefore, to be an empirical and philosophical jumping off point to explore some of those questions. It may well be in the future that our causal models of social stereotypes are so advanced that we can dispense entirely of blunt conceptual instruments like narrow and wide reach. But for now, this distinction provides us with a useful way to get the ball rolling.

3. "You're my doctor?": The Empirical Case for Wide Reach

Can stereotypes reach widely? With collaborators Jorge Morales and Chaz Firestone, I designed a set of experiments which put this question to this test. We choose to investigate

the doctor/nurse gender stereotype, because it is well known and widely held (remember the riddle!). We choose a *completely arbitrary* judgment, entirely unrelated to the stereotype: whether a person was facing left or right. This judgment is basic, perceptual, (fairly) low-level, and conceptually far-removed from stereotypes about doctors and nurses. Our experiments taught participants a simple regularity: doctors face one direction and nurses faced the other. We wanted to see if gender stereotypes about doctors and nurses inhibited participants' ability to learn and apply this new regularity, thereby demonstrating wide causal reach. In other words, are stereotypes so strong that if we taught participants an arbitrary rule doctors and nurses would they think we were teaching them about men and women?

3.1 The data

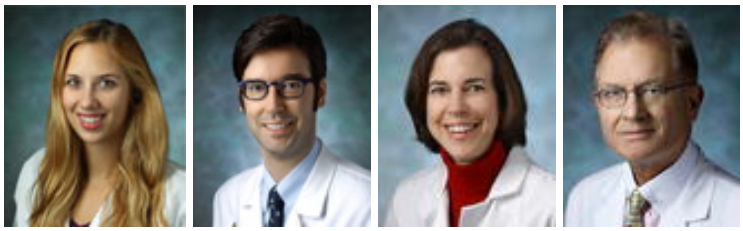

Figure 2. Example Doctor/Nurse Stimuli. Above are four examples of the physician headshots we used. Each participant saw 30 male headshots, half of which were randomly labeled as "doctors" and half as "nurses" and vice versa for the 30 female headshots. All doctors would appear facing the same direction and all nurses would appear facing the other (counterbalenced across participants).

We collected 60 standardized images (80x100px) of physicians from a major medical institution, half women and half men (see Fig. 2 above). 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/NURSE's]** shoulders?" appeared for two seconds above an empty frame, before a headshot appeared (see Fig. 3 below). Participants then indicated via a keypress whether the shoulders of the headshot subject were facing left or right. After the keypress, participants were asked to recall the

headshot subject's profession. So, participants just had to do two things during each trial: (1) indicate what direction the headshot subject was facing and (2) remember if the headshot subject was a doctor or a nurse. Unsurprisingly (given the simplicity of the task), we found that they were able to do both with near ceiling accuracy. Crucially, we introduced a simple learnable regularity by manipulating the headshots subjects' 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.

This meant that once participants learned this profession/orientation regularity, they would know what direction headshot subjects would be facing a full two seconds before the headshots even appeared (since they saw the "DOCTOR" or "NURSE" label two seconds before they saw the headshot). But the gender of headshot subjects was counterbalanced across "doctors" and "nurses"—so unlike profession label (which was *entirely predictive* of facing direction), the gender was *entirely unpredictive* of facing direction. Nonetheless, we wanted to see if participants' stereotypes about the gender of doctors and nurses would impact their ability to apply this simple and entirely predictive statistically learned regularity. In other words, because participants have the stereotype 'doctors are men', would it be harder for them to apply the regularity 'doctors face left' when judging the facing direction of female doctors?

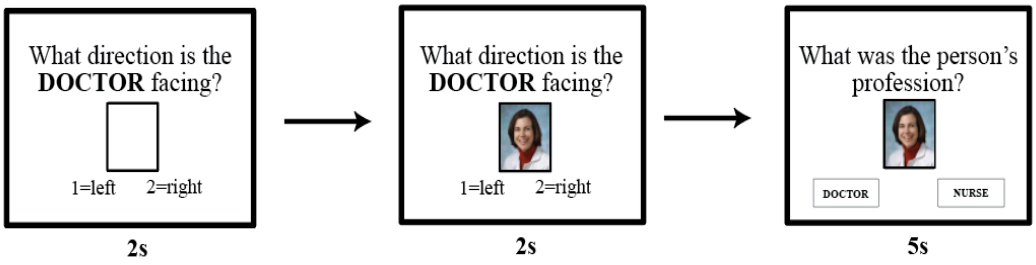

Figure 3. Experimental Design. Participants saw the question with the "DOCTOR" or "NURSE" profession label for two seconds before the headshot appeared. They then had 2

seconds to indicate the left or right orientation of the headshot subject's shoulders via a keypress. After they made the keypress, they were then asked to remember the headshot subject's profession.

In experiment 1 (with 100 participants, before exclusions) 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"). That is, even though the facing direction of the headshot was predicted only by the labeled profession and never by gender, participants were *still* slower to judge the facing direction female doctors and male nurses. Experiment 2 directly replicated this result with a larger sample (300 participants, before exclusions) and found similarly robust results. The reaction time differences are graphed below.[6]

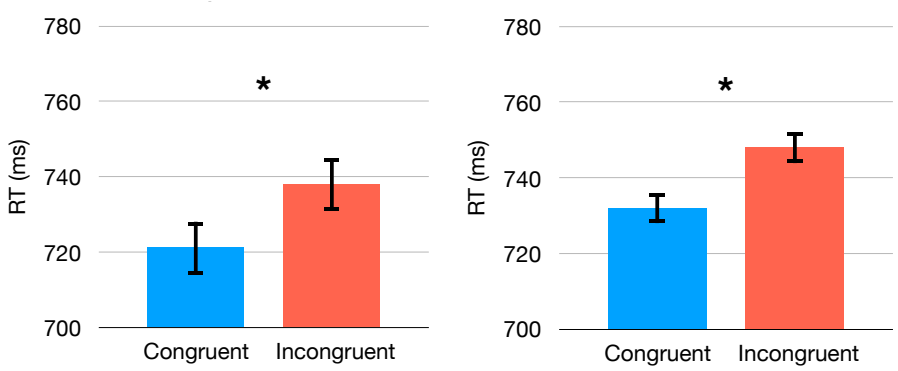

Figure 4. Reaction times graphed. In experiment 1 (left) participants were faster to judge the left/right facing direction in congruent (male "doctor" and female "nurse") trials than incongruent (female "doctor" and male "nurse) trials: 721ms vs. 738ms, t(72)=2.47, p=.016. In experiment 2 (right) we directly replicated this result: 732ms vs. 748ms, t(198)=3.72, p<.001.

[6] One thing to note when considering the relative reaction time differences between stereotype-congruent and stereotype-incongruent trials is the sheer simplicity of the task. While it's true that the reaction time difference between congruent and incongruent trials was only 17ms (which might sound small), note the task itself was *extremely easy*—the left/right judgments were straightforward, the headshots were not noisy, and headshot subjects were all facing a salient direction. Thus, what really matters for our purposes is the relative difference *between* the reaction time bars. When we are talking about an overall reaction time of only 721ms, a 17ms delay is huge. Moreover, it's especially telling that we were able to replicate the effect in experiment 2 with a very large sample size and get the exact same robust and statistically significant result. If the task were noisier or more difficult, we would expect to see the relative difference between the stereotype-congruent and stereotype-incongruent bars stay the *same* but the overall reaction times to go up.

This tells us that stereotype congruence is affecting participants' left/right orientation judgments (hereafter, 'the congruency effect'). But is statistical learning the mechanism of this interference? One might wonder if perhaps the effect was not being driven by learning the profession/orientation regularity at all but merely by brute surprisal—expecting to see a doctor and then seeing a woman or expecting to see a nurse and then seeing a man. To control for this, in experiment 3 we repeated experiment 2 with every aspect of the design held constant except for the regularity between profession and facing direction. So in experiment 3 (with 300 participants, before exclusions) each headshot (regardless of profession label) would be randomly assigned a facing direction. But without the profession/orientation regularity, we found that the reaction time difference between stereotype-congruent and stereotype-incongruent trials *completely* disappears: 769ms vs. 769ms. Experiment 3, thus, gives us a rare glimpse into the cognitive architecture of social stereotypes. Not only do we know our participants' gender stereotypes were impairing their ability to make spatial orientation judgments, we know this interaction was being facilitated by a statistically learned regularity—and we know this because the only difference between experiments 1 and 2 (where the effect robustly manifests) and experiment 3 (where the effect completely disappears) is the existence of the profession/orientation regularity.

So what exactly is happening here? We know the casual interaction between the stereotype and orientation judgment is being facilitated by the statistically learned profession/orientation regularity. But why does this occur?

3.2 Explaining the data

There seem to be two cognitive entities doing causal work here (where 'entity' just means some kind of implicitly or explicitly held belief or association). The first, which I will call 'R1' ('R' for 'rule' or 'regularity'), is just the doctor/nurse gender stereotype—doctors are men and nurses are

women. The second, which I will call 'R2', participants learn during the course of our experiment—doctors face one direction (say, left) and nurses face the other direction (say, right). Again, we know participants implicitly or explicitly learn R2 because the only difference in experimental design between experiments 1 and 2 (where the effect robustly manifests) and experiment 3 (where the effect does not manifest at all) is the introduction of R2.

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

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 two 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, enabling a fast orientation judgment response. However, in stereotype-incongruent "doctor" trials, where participants expected to see a left facing *man* and instead to saw a left facing *woman*, participants' expectations would be violated, creating a kind of cognitive lag which could explain the reaction time delay.

We suspect the congruency effect was also (at least in part) driven by participants actually forming a third rule, which I will call 'R3', that is entailed by the conjunction of R1 and R2.

R3. Men face one direction (the direction they learn doctors face in R2) and women face the other direction (the direction they learn nurses face in R2).

In other words, if participants learned that doctors face left and nurses face right (R2), then they might have inferred that *men* face left and *women* face right (R3), because they have the stereotype doctors are men and nurses are women (R1). We hypothesize that if participants came to believe R3—that men faced the doctor direction (say, left) and women faced the nurse direction (say, right)—, then seeing a left facing woman (or right facing man) would violate their R3 expectations about the facing direction of men and women. This violation could also account for the slower reaction times we see in stereotype-incongruent trials. Note here that even though R3 is entailed by R1 and R2, R3 does not hold true in our experiments (in fact, 50% of men in our experiments face left and 50% of men face right).

If the R3 interpretation is correct, then participants imported their gender stereotypes to completely invent regularities that were never there, which as a result impaired their orientation judgments. We currently are in the process of pre-registering new experiments which we hope will help us test the R3 interpretation—see chapter 2, experiment 4.

But readers at this juncture might be thinking: "yes, of course they would come to implicitly or explicitly believe R3 because it is entailed by the conjunction of R1 and R2—this is exactly what we should expect!" But in virtue of the logical relationship between R1 and R2 is it *really* so obvious that participants would deductively infer R3? I think not. For one, a lot of evidence suggests people are far from logically omniscient and often fail to recognize even basic logical entailments, especially if they don't explicitly believe one or more of the premises—as might be the case with the (perhaps implicit) doctor/nurse gender stereotype (Evans, Barston, & Pollard 1983; Evans 2006; Howarth, Handley, & Walsh 2016). Moreover, we often fail to appreciate logical

relationships because we do not bring all relevant premises to bear at the same time. In the philosophical tradition, fragmentationalist accounts of mind have tried to accommodate for this, stipulating that certain beliefs are available during some tasks and in some elicitation conditions and not in others (Lewis 1982; Stalnaker 1984; Egan 2008, Elga & Rayo ms; Mandelbaum 2016; Quilty-Dunn & Mandelbaum 2017).[7] The type of phenomenon has also been pointed to in the bounded rationality literature—given our cognitive limitations, not all our information can be available to us at all times for all tasks, which as an empirical matter of fact limits our rational capacities (Simon 1957; Kahneman 2003). The bottom line here is that our cognitive resources are limited and only a very small subset of our total information is made cognitively available to us—and generally this information is so available because it is relevant in some way to the cognitive task at hand (if I'm judging a dog show, dog show information would be made available and if I'm doing a math problem, math information would be made available).

Returning now to our experiments, because gender was irrelevant to the left/right orientation of headshot subjects, it should strike us as surprising that gender information would have been made cognitively available to participants *at all*. Hence, while it's true that the conjunction of R1 and R2 entails R3, it's not obvious that R1 information would be brought to bear during the learning of R2 in the first place.[8] Nonetheless, despite our fragmented natures and

---

[7] Egan (2008) cites 'failure-to-bring-to-bear' cases in support fragmentationalism like the following (p.10):

> "When I draw a blank in response to "what was Val Kilmer's character's callsign?" and respond with a confident "yes" to "was Val Kilmer's character's callsign 'Iceman'?", what credence, exactly, should we say that I assign to the proposition *that Val Kilmer's character's callsign was 'Iceman'*? There seems to be no happy answer to give – I'm disposed to act in some circumstances and in some respects like someone with a very high credence, and in other circumstances and other respects like someone with a much lower credence.

[8] Our participants already believe 'Fs are Gs' (stereotype) and then learn during our experiment 'Fs do H', which causes them to deductively infer 'Gs do H'. It would be interesting in future empirical projects to see if this pattern of inference holds for *non*-social properties. For example, if participants were taught to associate circles with being blue and then learned that blue things appeared on the left, would they come to believe circles

computation limitations, we find that R1 stereotype information *is* brought to bear when participants learn and apply R2! This result suggests that our gender stereotypes are so closely intertwined with our doctor and nurse concepts, that stereotype information will obligatorily be brought forth when doctor and nurse concepts are invoked (taking up our already limited cognitive resources), even if gender information is completely irrelevant to the task and impairs task performance.

The (distressing but significant) takeaway: thanks to your gender stereotypes, when you learn about doctors and nurses you can't *help* but think you're learning about men and women. Stereotypes are just *that* strong and infectious.

3.3 Putting it all together: Doctors, nurses, and wide causal reach

Now returning to the question of reach: can stereotypes reach widely, impacting judgments not directly related to the stereotype's content? Yes! To recap, while we stereotype doctors as being men, we don't have preconceived notions of what direction they face in professional headshots. Nonetheless, our experiments clearly demonstrate that participants' stereotypes were impairing their left/right orientation judgments. And we know (because the congruency effect manifested in experiments 1 and 2 but not in control experiment 3) that wide causal reach is being facilitated by participants learning the regularity that doctors face left and nurses face right (even though the this regularity has nothing to do with gender!), causing them to falsely come to believe that *men* face left and *women* face right. This suggests that stereotypes are *so* strong that they exert causal influence on nearby learned associations, even if the nearby

appeared on the left? Would color intrude on orientation judgments in the same way gender intruded on orientation judgments in our experiments? If asked to speculate, I would suspect that we would *not* see the same patterns of inference because stereotypes—unlike shape/color associations—are learned very young and are constantly reinforced by external social structures. Hence, I would guess that it is the *strength* of social stereotypes compelling participants to form the R3 inference (indeed, it would be difficult to reinforce 'circles are blue' to the same degree that culture reinforces 'doctors are men'). However, if Fs and Gs were non-social properties but were reinforced to the same degree, then we might expect see similar patterns of interference.

associations have nothing to do with the content of the stereotype. Thus, our results demonstrate an instance of wide causal reach (showing that wide reach *is* possible!) and point to a mechanism for wide causal reach—anchoring via statistical learning of an intermediate regularity. These findings tell us something both novel and surprising about the cognitive architecture of social stereotypes.

But, of course, we don't in fact learn to associate doctors and nurses with left/right facing direction. So perhaps an ecological validity worry crops up here. What does wide causal reach matter for how people actually go about forming beliefs and making judgments?

We designed our experiments to see if *low-level*, *arbitrary* judgments could be impacted by social stereotypes—which should be the *hardest* case for wide reach. Why is this the hardest case? First, one might be more likely to expect wide causal reach if the task took longer or was more difficult because participants would have more time to go through enough inferential steps to eventually activate the social stereotype. However, we ensured that the task itself was quick (on average participants responded in well under a second) and easy (accuracy was near ceiling). Second, one might guess that if stereotypes could widely reach to these sorts of fast, arbitrary judgments at all they would only be able to do so following extensive training trials. However, our participants only saw six training trials (which we gave them to ensure they knew how to record their keypress responses correctly). Nonetheless, over the course of a few minutes their fast, low-level, perceptual judgments were significantly impaired by their social stereotypes.[9]

But if the 'hard case' arbitrary orientation judgments could be impaired by social stereotypes, then surely any number of judgments are potentially susceptible to the same type

[9] Moreover, our data suggests the effect starts taking place almost immediately. There were not significant reaction time differences in the first and second half of the trials, evidencing the quickness and ease of this type of statistical learning.

of interference. Consider a slightly nearer judgment to the doctor/nurse gender stereotype: what kind of car someone is driving. Car make is certainly not a part of the content of the doctor/nurse gender stereotype, but it is still closer to the stereotype than left/right facing direction—car make is associated with socio-economic status, which is associated with male dominated professions like medicine. So, if I learned that doctors tend to drive BMWs would it be more difficult to recognize that my female doctor friend's new car is a BMW if I saw her drive up? The arbitrariness of the profession/orientation regularity suggests 'yes'. We taught our participants the profession/orientation regularity precisely *because* it is so arbitrary. Thus, if even extremely arbitrary wide reaching judgments (like facing direction) can be affected by social stereotypes, then we should certainly expect that less arbitrary wide reaching judgments (like the kinds of cars people drive) could also be affected.

Of course, we have many social stereotypes and—whether we are always explicitly aware of it or not—we are constantly in the process of identifying and detecting regularities to help us efficiently interact with our environments (doing so is necessary—think back to the fire alarm case!). So if stereotypes can causally influence judgments *any time* a new statistical regularity between a quality associated with a stereotype (e.g. being a doctor) and an attribute unrelated to the stereotype (e.g. facing left, driving a BMW, etc.) is learned (no matter how arbitrary that regularity is!), then stereotypes can potentially impact many different kinds of wide reaching judgments. This is scary indeed!

But, does wide reach really matter in a larger sense? One of the reasons we study social bias at all is to understand why people engage in discriminatory patterns of behavior towards negatively stereotyped social groups. And while understanding the cognitive architecture of social stereotypes can help us conceptualize the role bias plays in our mental lives, one might wonder if (and how) wide causal reach informs larger patterns of social discrimination, which

we see stereotyping as being connected to. I will conclude by gesturing towards one possible answer: wide reach translates into considerably more processing *dis*fluency, which leads to more prejudice and discrimination.

4. Processing Fluency and Discrimination

Alter and Oppenheimer (2009) characterize processing fluency as "the subjective experience of ease with which people process information", arguing it is a "metacognitive cue that plays an important role in human judgment" (219). So, processing fluency involves the difference in *phenomenology* that accompanies relative ease of cognitive processing.[10] For example, think about the phenomenological difference between doing a simple math problem and doing a difficult math problem. Simple math problems just *feel* easier to cognitively process, which we take as a cue that our answers are more likely to be right. It has been demonstrated that we use processing fluency as a cognitive heuristic for judging truth, preferability, and trustworthiness—if something feels easier to cognitively process, we tend to think it must be more likely to be true, fitting, or better.

For example, in one of the first experiments on the effects of processing fluency, Schwatz et al. (1991) examined assertiveness judgments. Participants were asked to either recall 6 or 12 examples of their own assertive behavior, after which they rated their own assertiveness. People who only had to recall 6 examples of assertive behaviors rated themselves as more assertive than those who were asked to recall 12 examples. The researchers argued

[10] Note that processing fluency is not exactly the same as perceptual fluency (although types of perceptual fluency might be instances of processing fluency). Perceptual fluency refers to how quickly you are able to *perceptually* process a scene. We know from control experiment 3 that the congruency effect *isn't* merely being driven by difficulty perceptually processing a woman as a doctor or man as a nurse. But classic examples of processing fluency involve how fluently people are able to make judgments. This is the type of pressingly fluency which will be most relevant to the present discussion of wide reach. Thanks to Judy Kim and Steven Gross for pushing me on this.

that participants were using their experience of fluency (or disfluency) as evidence for how assertive they actually were. Because recalling 12 examples of assertive behavior was a more difficult task than recalling 6, those that had to recall 12 judged themselves to be less assertive. In work that followed, manipulations of processing fluency were claimed to influence a wide range of judgments. Regarding truth judgments and visual cues, it's been argued that statements written in easier to read colors (Reber & Schwarz 1999) and fonts (Alter et al. 2007) were judged more likely to be true. In the domain of consumer choice, simpler product selections (i.e. having fewer products to choose from) were judged to be preferable to more difficult product selections (i.e. having more products to choose from) (Iyengar & Lepper 2000). And regarding person-level judgments, Oppenheimer (2006) found that using longer, obscure words caused readers to judge the author to be less intelligent. Thus, more fluently processed statements, decisions, or pieces of writing are judged as more preferable and likely to be true.

Going back to social stereotypes, we know people are slower to judge the facing direction of stereotype-incongruent headshots. But are judgments about stereotype-incongruent headshots also felt to be more *disfluent*? In other words, is the reaction time difference between stereotype-congruent and incongruent trials accompanied by a difference in subjective experience, suggesting that participants' stereotype-incongruent orientation judgments are less fluently processed? The phenomenology of taking the experiment is quite powerful. To quote one participant who left a comment at the end of the experiment, "the people that look like they would be doctors are not always, and the ones you definitely think are nurses are actually doctors". Thus, stereotype-incongruent trials take longer and *feel* more difficult to participants, which suggests that stereotype-incongruent trials *are* being processed more disfluently. So we have reason to think that wide reach can cause disfluency in cognitive

processing. This means that because many judgments can be potentially impacted by wide reaching social stereotypes (as I argued in the last section), the existence of wide causal reach creates the potential for vast processing disfluency.

But how does fluency relate to *discrimination*? Well, considering the relationship between processing fluency and truth and preferability judgments, I think we should assume that the *more* judgments (narrow or wide) about a person that are disfluent, the *less* reliable and trustworthy they will seem. This means that social groups associated with more disfluency will seem less fitting in their respective social roles. For example, experiencing disfluency when making even basic arbitrary judgments about a female doctor (i.e. wide reaching judgments) might make her seem more "off" and less believable in her role as a doctor than her male counterparts. This could explain why many patients still prefer to see male doctors—engaging with male doctors will be experienced as more fluent, which the fluency heuristic falsely causes us to interpret as male doctors *in fact* being more competent. Consequently, people engage in discriminatory actions towards members of negatively stereotyped groups—for example, not seeing a female doctor—because they associate interacting with members of those groups as feeling disfluent and thus uncomfortable. In this way wide causal reach means more disfluency which translates into more discrimination.

I want to conclude by suggesting that this explanation can tell us something novel about the causal structure of *implicit* bias. We seem to have some grasp on the causal relationship between *explicit* bias and discrimination—if you are explicitly biased against a social group you'll be motivated to discrimination against members of that group. However, while it is assumed that implicit bias can motivate discrimination, it is not always clear what this causal relationship looks like. How would stereotypes you don't explicitly endorse compel you—without your knowledge or awareness—to partake in discriminatory actions? Where

does the bias get in? The empirical and philosophical literature offers few answers. But we now have one possible answer: wide causal reach creates more processing disfluency and more processing disfluency translates into more discrimination.

Chapter 2:

"You're my Doctor?":

Stereotype Incongruent Identities Impair Recognition of Incidental Visual Features

Austin A. Baker[1, 2], Jorge Morales[2], and Chaz Firestone[2]

[1]Department of Philosophy, Rutgers University, New Brunswick
[2]Department of Psychological and Brain Sciences, Johns Hopkins University

1. Abstract

Social stereotypes shape our judgments about people around us. What types of judgments are susceptible to such biased interference? A striking example of stereotype bias involves the treatment of people whose identities run counter to our stereotypes—as when women are assumed to be students, research assistants, or nurses rather than professors, principal investigators, or doctors. But can such stereotypes also intrude on representations that have nothing to do with the content of the stereotype in question? Here, we explore how the assumptions we make about other people can impair our ability to process completely incidental, and surprisingly low-level, aspects of their appearance—including even their location in space. In our experiments, participants learned to associate "doctors" and "nurses" with left or right facing directions. Surprisingly we find that participants' gender stereotypes impaired learning of this basic perceptual rule. Thus, even straightforward forms of statistical learning (here, between profession labels and spatial orientations) can be intruded upon by long-held social biases.

2. Statistical Learning

We choose to test the causal efficiency of stereotypes by investigating the relationship between social stereotypes and the statistical learning of low-level perceptual features. Statistical learning as a concept was first introduced to describe infants' sensitivity to the co-occurrence to certain patterns of syllables, enabling them to extract segmentation of words from fluent speech (Saffran, Aslin, & Newport 1996; Aslin, Saffran, & Newport 1998). It has also been claimed to be involved in category learning (Orbin et al 2008), serial reaction time tasks (Misyaki & Christiansen 2012), conditioning (Courville, Daw, & Touretzky 2006), visuamotor learning (Hunt & Aslin 2001), and visual search (Baker, Olson, & Behrmann 2004). Given the strength and ubiquity of statistical learning across numerous domains, we wondered—as a test of the causal power of social stereotypes—if stereotypes could impair even perceptual statistical learning.

To explore this question, we choose the stereotype 'doctors are men and nurses are women'. We selected an arbitrary low-level perceptual regularity which our participants would statistically learn—that doctors face one direction and nurses face the other direction (either left or right, counterbalanced across participants). We wondered if participants' gender stereotypes would intrude upon their statistical learning of this regularity such that if they were taught '*doctors* face left' they would instead come to learn that '*men* face left'.

3. Experiments 1 & 2

3.1 Participants

In experiment 1 we recruited 100 participants (before exclusions) and in experiment 2 we recruited and 300 participants (before exclusions) through Amazon Mechanical Turk. For more on the fidelity of Amazon Turk on cognitive behavioral experiments see (Crump,

McDonnell, & Gureckis, 2013). The size, hypothesis, and exclusion criteria of these studies were preregistered.

3.2 Stimuli and procedure

We collected 60 standardized images (80x100px) of physicians from Johns Hopkins Hospital—half men and half women. All headshot participants were wearing lab coats without discriminable writing. Each headshot had a salient facing direction (either left or right; normed in a separate study) that we could manipulate in advance by flipping the image. Participants were told that the purpose of the experiment was to see how quickly and accurately they could answer simple questions about medical professional's headshots. Gender, stereotypes, and bias were never at any time mentioned or made salient.

On each trial, the question "What's the direction of the **[DOCTOR/NURSE's]** shoulders?" was shown for two seconds above an empty frame, before a headshot appeared. Participants then had to indicate via a keypress whether the shoulders of the person in the headshot were facing left or right. After their keypress response was recorded, the question with the "DOCTOR" or "NURSE" labeled disappeared and participants were asked (while the headshot was still visible) if the person was a doctor or a nurse. Performance feedback was given throughout. Participants who took longer than 2 seconds answering the left/right orientation question or longer than 5 seconds answering the doctor/nurse profession label question were given "too slow" feedback and were prevented from responding. The first six trials were training trials—the data of which we did not collect—to ensure participants understood the instructions and knew how to record their keypress selections correctly.

Two exclusion criteria were preregistered. First, we only analyzed trials where participants answered both the facing direction question and the profession label question

correctly within the specified 2 and 5 second time constraints. Second, we did not analyze the data of any participant who did not have an overall accuracy of 80% across all trials.

For each participant, all 60 headshots were assigned a "doctor" or "nurse" profession label and left or right facing direction orientation. Half of the 30 female headshots were randomly labeled as "doctor" and half labeled as "nurse" and half of the 30 male headshots were randomly labeled as "doctor" and half labeled as "nurse". Therefore, women were equally likely to be labeled as "doctors" or "nurses" and equally likely to appear facing left or right. Each participant was also randomly assigned to one of the two possible regularity conditions: (1) either doctors faced left and nurses faced right or (2) doctors faced right and nurses faced left. This meant that once participants learned the regularity (hereafter, the 'profession/orientation regularity'), they would be able to know two seconds before the headshot appeared what direction the headshot subject would be facing based on their profession label.

3.3 Results

To see if participants' gender stereotypes were impairing their ability to apply the profession/orientation regularity, we split participant trials into two within-subject conditions: a stereotype-congruent condition (male "doctors" and female "nurses" trials) and a stereotype-incongruent condition (female "doctors" and male "nurses" trials). We ran a two-tailed paired-samples t-test between the congruent and incongruent conditions and found that participants were slower to judge the orientation of stereotype-incongruent headshots (female "doctors" and male "nurses", mean reaction time 738ms) than stereotype-congruent headshots (male "doctors" and female "nurses", mean reaction time 721ms); $t(72)=2.47$, $p=.016$. We directly replicated these results with a larger sample size in experiment 2—748ms vs. 732ms; $t(198)=3.72$, $p<.001$.

4. Experiment 3: Controlling for Surprise

But one might wonder if perhaps the congruency effect had nothing to do with learning the profession/orientation regularity, but rather (e.g.) the oddness of expecting a "nurse" and then seeing a man, which might impair judgments for independent reasons. Is the congruency effect is being driven by sheer surprisingness or are gender stereotypes genuinely impairing participants' statistical learning? We ran experiment 3 to control for this alternative possibility.

4.1 Participants

As in experiment 2, in experiment 3 we recruited 300 participants (before exclusions) through Amazon Mechanical Turk. The size, hypothesis, and exclusion criteria of this study were preregistered.

4.2 Stimuli and procedure

We repeated the experiment 2 with every aspect of the design held constant except for the regularity between profession and facing direction. Participants in experiment 3 were not assigned to a regularity condition—each headshot (regardless of "doctor" or "nurse" profession label) was randomly assigned a left or right facing direction.

4.3 Results

Like experiments 1 and 2, participant trials were split into two within-subject conditions: a stereotype-congruent condition (male "doctors" and female "nurses" trials) and a stereotype-incongruent condition (female "doctors" and male "nurses" trials). We ran a two-tailed paired-samples t-test between the congruent and incongruent conditions and found that without the profession/orientation regularity (i.e. when participants are not assigned to a regularity condition), the reaction time difference between stereotype-congruent and stereotype-incongruent trials *completely* disappears—the mean stereotype-congruent and stereotype-incongruent reaction times

were the same: 769ms vs 769ms. Because the only difference between experiments 1 and 2 and experiment 3 was the assignment of a regularity condition, we know then that the reaction time differences were being driven by participants in experiments 1 and 2 statistically learning the profession/orientation regularity.

5. Discussion and Further Directions

We suggest that gender intruded into the learned regularity between profession and orientation, even though the "profession" labels were completely arbitrary. In other words, our experimental design taught participants the rule that *doctors face left and nurses face right*; then, participants inferred on their own that it must also be true that *men face left and women face right*, because they harbor the social stereotype that doctors are men and nurses are women. If this interpretation is correct then participants imported their own biases to completely invent non-existent regularities, leading to the reaction time difference between stereotype-congruent and stereotype-incongruent trials that we observe in experiments 1 and 2. We are in the process of preregistering experiment 4, which tests the 'invented regularity' hypothesis.

5.1 Experiment 4

In experiment 4 we will recruit 600 participants from Amazon Mechanical Turk. The participants will be split into two groups. The first 300 of the participants (the 'regularity group') will complete the same task as participants in experiments 1 and 2 and will be randomly assigned to one of the two possible regularity conditions: (1) doctors face left and nurses face right or (2) doctors face right and nurses face left. They will then complete the 60 trials. The last 300 participants (the 'non-regularity group') will complete the same task as participants in control experiment 3. They will not be assigned to a regularity condition and facing direction will be randomly assigned for each headshot.

First, we will perform an ANOVA on the reaction times of participants' left/right orientation judgments to ensure we find the same effects we observed in experiments 1, 2, and, 3, testing for the main effect of the profession/orientation regularity, the main effect of stereotype congruency and incongruency on the reaction times of participants' orientation judgments, and most importantly, the interaction between stereotype congruency and regularity presence. If we find a significant interaction, we will run two post hoc t-tests for the reaction time difference between stereotype-congruent and stereotype-incongruent trials in each (the regularity and the non-regularity) group.

After the participants in both groups complete the 60 trials, they will answer a series of multiple-choice questions about the experiment. Note: we will only analyze two of the multiple-choice questions; all others will be distractors which obscure the purpose of the questions.

5.2 'Invented' regularity

The first multiple choice question we will analyze is the following:

In the experiment:

A. MEN tended to face LEFT and WOMEN tended to face RIGHT

B. MEN tended to face RIGHT and WOMEN tended to face LEFT

C. Neither/don't know

We want to see if participants in the regularity group select the option which indicates that they came to believe that *men* face the same way they learned *doctors* face (and vice versa for women and nurses). For example, after being exposed to 30 left facing doctors (half or whom were men and half of whom were women) would participants walk away from the experiment

thinking that men tended to face left? If we found that they did, this would provide us with solid evidence that participants' gender stereotypes caused them to explicitly *invent* a new regularity (between gender and left/right spatial orientation) on the basis of the regularity they statistically learned (between profession and left/right orientation).

To test this, we will run a chi-squared test comparing the proportion of participants in the regularity and non-regularity groups who invented 'men face one direction women face the other direction' regularity. Specifically, if our hypothesis is correct and participants in the regularity group who are assigned the 'doctors face left and nurses face right' regularity invent the 'men face left and women face right' regularity, then the proportion of regularity group participants who select option A over option B should be higher than the proportion of non-regularity group participants who select option A over option B. And vice versa for participants who are assigned to the 'doctors face right and nurses face left' regularity condition—the proportion of regularity group participants who select option B over option A should be higher than the proportion of participants in the non-regularity group who select option B over option A. We further predict that if regularity group participants are inventing the 'men face one direction women face the other direction' regularity, we should also see less people selecting option C in the regularity group than in the non-regularity group.

5.3 Secondary analyses

In response to our first three experiments, people have expressed curiosity in how participants learned the profession/orientation regularity. As they went through the trials, were participants *explicitly* aware that all the doctors were facing one direction and all the nurses were facing the other or was this regularity *implicitly* learned and applied? Thus, as a secondary analysis, we want to determine the extent to which participants are explicitly learning the profession/orientation regularity. To test this, we will also ask participants the following

question to see if there were explicitly aware they were learning the profession/orientation regularity.

In the experiment:

A. DOCTORS tended to face LEFT and NURSES tended to face RIGHT

B. DOCTORS tended to face RIGHT and NURSES tended to face LEFT

C. Neither/don't know

We will run a chi-squared test comparing the proportion of participants in the regularity and non-regularity groups who explicitly learned the profession/orientation regularity. For participants in the regularity group who are assigned the 'doctors face left and nurses face right' regularity, if the proportion of regularity group participants who select option A over option B was higher than the proportion of non-regularity group participants who select option A over option B, then we know the regularity is being explicitly learned. And vice versa for participants who are assigned to the 'doctors face right and nurses face left' regularity condition—if the proportion of regularity group participants who select option B over option A is higher than the proportion of participants in the non-regularity group who select option B over option A, then we know the regularity is being explicitly learned.

We are also interested in investigating the relationship between explicit learning of the regularity and the reaction time effect we observe in experiments 1 and 2. In particular, do people that explicitly learn the regularity exhibit the effect to a greater degree than people that don't explicit learn it? To test this, we will split the data from the regularity group between participants who explicitly learn the regularity and participants who did not, to see whether explicitly learning the regularity drives the reaction time differences between stereotype-

congruent and stereotype-incongruent trails. We will then perform an ANOVA on the main effect of having explicitly learned the profession/orientation regularity, the main effect of the reaction time difference between stereotype-congruent and incongruent trials, and, most importantly, the interaction between explicitly learning the profession/orientation regularity and the reaction time difference between stereotype-congruent and incongruent trials. If we find a significant interaction, we will run two post hoc t-tests for the reaction time difference between stereotype-congruent and stereotype-incongruent trials for regularity group participants that explicitly learned the regularity and regularity group participants that did not explicitly learn the regularity. This should show us if there is an interaction between explicit learning of the regularity and the congruency effect.

We will also collect demographic data (gender, age, and education level) after the survey questions to see if there is any relationship between the congruency effect and participants' gender, age, or education level.

5.4 Further experiments

Lastly, we have plans to run a version of these experiments with race, showing participants pictures of Caucasian and African American men in collared shirts labeled either as "professors" or "waiters" and asking what direction the men subjects were facing. We are interested to see if participants' stereotypes about race—like their stereotypes about gender—interact with their spatial orientation judgments. This will help us understand how universal the congruency effect is.

Chapter 3:

Accuracy vs Action Guidance:

Attention and the Function of Perception

1. Introduction

What is the function of perception? In philosophy and cognitive science, it is widely assumed that perception functions to produce perceptual states that accurately represent the world to us. D. H. Mellor's characterization of perception speaks to this idea: "[P]erceptual experience isn't like fear or desire. Like belief, although it may be false, it seems to aim solely at truth: it's function, after all, is to tell us how the world truly is" (1988, 149, my emphasis). Perception, on this account, functions to furnish us with an accurate picture of the world. However, I will argue that the view that accurate representation is the one and only function of perception (hereafter, the 'accuracy view') ignores the distinct and important action guiding function of perception. My aim here will be to specifically consider the function of perception in light of a mounting body of empirical literature on the effects of attention on perceptual phenomenology. Thus, in this paper, I will argue that:

(i) perception functions to *guide action* that promotes differential reproduction (i.e. reproduction in the virtue of an organism's adaptation to its environment), which I will hereafter call the 'the action guidance view'

(ii) the action guidance view does not end up collapsing into the accuracy view (i.e. accurate perceptual states will not always be the ones that guide action in a way that promotes differential reproduction), and

(iii) the large body of recent research on the effects of voluntary (endogenous) and involuntary (exogenous) attention on perceptual phenomenology is better explained in terms of the perceptual system functioning to produce action guiding perceptual states rather than accurate perceptual states. In section 4, I will discuss two such attention studies popularized in the empirical and philosophical literature (Carrasco, Ling, & Read, 2004 and Liu, Abrams, & Carrasco, 2009).

Finally, though I not deny that accurate representation might be *a* function of perception, I will stress the importance of action guidance as a—frequently overlooked—function of perception. Moreover, I argue that even if we accept that one of the functions of perception is to supply us with accurate representations, the empirical work I engage with suggests that there are certain perceptual phenomena that are mediated by the action guiding, rather than accuracy, function of perception. If my interpretation is correct, then perception is in fact not solely in the business of accuracy, and philosophers and cognitive scientists who restrict themselves to discussion of perceptual states in terms of the accuracy will be unable to satisfactorily explain a wide range of perceptual phenomena.

I want to flag one point of definition before diving in. I will largely assume the teleological view of function, according to which if *f* is the biological function of a system *S* of an organism *O*, then *S* was selected to preform *f* because preforming *f* contributes to *O*'s differential reproduction.[11] And while I think my argument could potentially be applied to

[11] It's worth nothing that teleological views have been interpreted in importantly different ways. Griffiths (1993) and Godfrey-Smith (1984) have argued that the selection mechanisms can be relatively recent, whereas Millikan (1989) and Neander (1991) have argued that the relevant sense of selection pertains to the *O*'s evolutionary history and the way system *S* performing *f* contributed to *O*'s ancestors survival. I will try to remain neutral between the two interpretations of the teleological view.

other theories of biological function as well, it certainly most naturally fits with teleological view (and it seems as if proponents of the accuracy view likewise assume some version of the teleological view).

2. Accuracy as the Function of Perception

The accuracy view, according to which the perceptual system functions to produce perceptual states that accurately represent the world, has been historically popular in philosophy and cognitive science. It also seems to be rooted in intuitive common sense. When I look out my bedroom window and see a taxi on the street below, I assume that I am accurately perceiving the street. From an intuitive perspective, what could perception function to do if not telling us how the world is? So, the accuracy view gets points for intuitive appeal. But how should we weigh intuitive plausibility when it comes to considering biological function?

It's worth noting that biological functions are not typically introspectively obvious. For example, the inclination to lick a wound might seem to be compulsive and meaningless. However, wound licking behavior has been observed in many different species and it was recently discovered that saliva contains histatins, which are microbial proteins that kill bacteria and promote accelerated healing. To the function of wound licking is now thought to be warding off infection and promoting healing. But, when we lick our cut finger, most of us are not in any sense aware that the action has this function—we just do it unreflectively. Hence, biological function is not always introspectively transparent, meaning that our unreflective commonsense assumptions about biological functions (like the function of perception) might end up being false.

But, introspective limitations aside, the accuracy view continues to dominate philosophy and cognitive science and has many champions. To understand the view, we must consider the ways it has been interpreted.[12]

2.1 Interpreting the accuracy view

Tyler Burge argues that the biological function of the perceptual system cannot be accuracy alone because biological function is connected to fitness and accuracy "in itself", he claims, does not contribute to fitness (2010, 301—my emphasis):

> Biological functions are functions that have ultimately to do with contributing to *fitness for evolutionary success*. Fitness is very clearly a practical value. It is a state that is ultimately grounded in benefit of its effects for survival for reproduction. Explanations that appeal to biological function are explanations of the practical (fitness) value of a trait or system. But accuracy is not *in itself* a practical value.

Indeed, accuracy would be a useless attribute of perceptual states if the states did not in *some* way contribute to the fitness of the perceiver. It is hard to see how a perceptual system could be selected for that produced very accurate perceptual states that in no way contribute to the organism's differential reproduction.

To illustrate the uselessness of accuracy "in itself" consider a hypothetical organism we will call 'Perceptor'. Perceptor cannot effectively interact with its environment (its perceptual system is not connected to its action system), but perceives everything with perfect accuracy. A predator approaches Perceptor. Though Perceptor's perceptual system produces a stunningly accurate perceptual state of the approaching predator, Perceptor lacks the capacity to act accordingly. The accuracy of the 'a predator is approaching' perceptual state does not

[12] The accuracy view is the largely orthodox position in cognitive science and has been defended widely in the literature over the last ten years. In addition to those cited in section 2.1, see also Trivers (2011), also explicitly defends a version of the accuracy view. Relatedly, Hoffman, Singh, and Prakash (2015) emphasize the connection between the accuracy view and Bayesian models of perception (though they ultimately reject the accuracy view), citing Pizlo, Sawada, and Steinman (2014), Yuille and Bülthoff (1996), and Geisler and Diehl (2003) as defenders of the accuracy view from the Bayesian perspective.

contribute to Perceptor's fitness and continued differential reproduction. Perceptor thus becomes lunch. Evolution is in the business of practicality, fitness, and differential reproduction; evolution, to quote philosopher Peter Graham, "does not care about veridicality" (2014b, 23). Hence, more accuracy does not *necessarily* amount to a fitness advantage.

Nonetheless, even if evolution does not select for accuracy *in itself*, defenders of the accuracy view can argue that accurate representation is the function of the perceptual system because having a perceptual system that produces accurate representations *in fact* enhances biological fitness and differential reproduction. So, even if accuracy for the sake of accuracy is useless, if it was the case that generally having accurate perceptual states enhances fitness, then accurate representation might have been selected for. Of course, this interpretation of the accuracy view assumes that having accurate representations enhances fitness in creatures *like us*. Peter Graham argues for such an accuracy view (2014b, 22—my emphasis):

> The accuracy of perceptual representations—especially visual representations in humans—plays a role in the functional analysis of how organisms with perceptual systems are able to survive and reproduce. *Getting it right often contributes to fitness, as a contingent, empirically determined matter of fact, in countless creatures with perceptual systems*. Just take away accuracy but leave everything else intact and see what happens. Would you rather walk towards a cliff with accurate, or inaccurate, representations as your guide?

Responding to Burge, Graham goes on to say the following about this espoused contingent evolutionary benefit of having accurate perceptual representation (2014b, 23—my emphasis):

> And so it does not follow from the fact that evolution does not care about veridicality per se that it does not care about veridicality as a contingent, empirically well-established matter of fact. All the point shows is that if accurate representations did not contribute to fitness, nature would not have cared about them. But *since they do*, nature cares.

Thus, the fitness value of accuracy leads Graham to conclude that "human perceptual systems—especially visual systems—have producing reliably accurate perceptual representations as a biological function" (2014b, 1; he argues similarly in 2010, 2012, and

2014a). According to Graham, accurate representation is the contingent function of the perception system *because it confers considerable differential reproduction benefit.*[13]

This interpretation of the accuracy view is popular outside of philosophy as well. Psychologist and vision scientist Stephen Palmer stresses the evolutionary role of accuracy, claiming that the "evolutionary role of visual perception is to provide an organism with accurate information about its environment" (1999, 15). Palmer asserts that the accuracy view construed in this way explains why so many of our perceptual experiences are accurate—the perceptual system functions to deliver accurate perceptual states and, thankfully for us (since he thinks accuracy enables survival and reproduction), it often fulfills this function.

Indeed, there is a tendency to defend the accuracy view by citing cases of perceptual accuracy ('look at how accurate perception is, it must function to be that way!'). This argument involves a kind of inference to the best explanation from the purported instances of perceptual accuracy. However, if we grant that many of our perceptual states are accurate it does not follow from 'we have a lot of accurate perceptual states' that 'accurate representation must be the function of the perception' because biological systems can have functions that they rarely ever preform. Millikan makes this point, saying (1989, 295):

> It is not always true that typical items falling in a function category perform that function. It is quite possible, for example, that the typical token of a mating display fails to attract a mate, and that the typical distraction display fails to distract the predator.

Therefore, frequency alone doesn't tell us anything about function. In this way, merely observing that many perceptual states are accurate does not necessarily support the accuracy

---

[13] The actual evolutionary benefit of veridical perception is quite contentious. I will in section 4 argue that certain kinds of perceptual inaccuracies are beneficial. In the last few years the topic of adaptive misrepresentation has become quite popular. Hoffman, Singh, and Prakash (2015) argue that perception is, in fact, not veridical and, by appealing to evolutionary game theory, they demonstrate how having veridical perception considerably hinders differential reproduction. So, like Burge, they argue that evolution does not select for veridicality and only selects for fitness and differential reproduction. But, unlike Graham, they assert that veridical representation does not *in fact* enhance fitness. For a critique of Hoffman, Singh, and Prakash see Cohen (2015).

view. Citing evidence in favor of theories of perceptual function will require more nuance.

Of course, we are often limited in what information we have about biological systems. So—stepping away from the accuracy view for a moment—how ought we actually, in a scientifically-informed and appropriately nuanced way, make determinations about the functions of biological systems? What information will be relevant for these kinds of determinations? The answer to this will probably differ depending on the system we are considering. However, for most biological systems, we can imagine it will be relevant to consider (1) the organization of the system, (2) the way the system performs actions, (3) the outputs of the system, and (4) the relationship between the system and biological fitness (particularly the biological fitness of the organism's ancestors).

For example, while it's true that the heart expands and contracts when it beats, we do not think the heart *functions* to expand and contract. When we consider the organization of the heart (the two ventricles, the mitral valve, the tricuspid valve, the aortic valve, and the pulmonary valve), the actions it performs (pumping blood, removing waste, expanding and contracting, making noise, etc), and the relationship between the heart and the biological fitness of the organism, it is clear that the heart functions to pump oxygenated blood around the body and remove waste. While expanding and contracting enables the heart to pump blood around the body, the heart does not *function* to expand and contract.

There is much more that could—and indeed should—be said about making informed, empirically supported, determinations about the function of biological systems. However, the four considerations put forth above will serve as a useful starting point for thinking biological function. Moreover, these holistic considerations should also help us get past 'the systems Φs to it must function to Φ' types of argument. My approach in this paper is meant to be both holistic and data driven; while I ultimately point to cases of perceptual inaccuracy and argue

that in these instances perception isn't functioning to accurately represent, I do so in a way that considers the organization of the perceptual system and the system's actions in light of the perceiver's biological fitness.

2.2 Perceptual inaccuracy vs. accuracy 'side effects'

So, the proponent of the accuracy view cannot merely point to cases of perceptual accuracy in support of their position. What, then, should they say perceptual *in*accuracy? Clearly in light of all we have said here, the existence of perceptual inaccuracy is not *by itself* evidence against the accuracy view (because, again, a system does not have to always fulfil its function). However, we will still want to know how any view of perception accounts for perceptual inaccuracy. Frequently cited examples of perceptual inaccuracy involve visual illusion, like the Müller-Lyer illusion (below).[14]

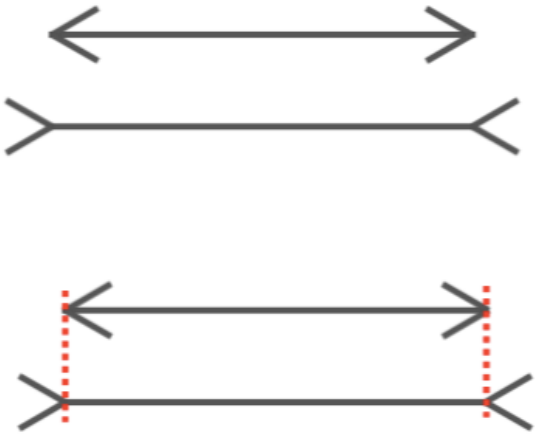

Figure 5. The Muller-Lyer Illusion

All four horizontal lines in Fig. 1 are the same length, even though we see the horizontal lines

[14] Note, I will discuss Müller-Lyer as a well-known example of a perceptual illusion to illustrate the notion of (what I will call) an accuracy 'side effect', but nothing I say about accuracy 'side effects' are particular to either the Müller -Lyer illusion or the size constancy scaling explanation of the illusion.

between the arrow heads as being shorter than the horizontal lines between the arrow tails. As is the case with many visual illusions, while we can know this (and even measure the lines to establish this fact for yourself) we cannot help but see the line between the arrow heads as being shorter than the line between the arrow tails. The standard explanation for the Müller-Lyer illusion involves appeal to size constancy scaling.[15] The perceptual system takes into account surrounding depth cues when scaling the size of an object to represent the object's size consistently (even though the object's size on the retina will change). In the case of the Müller-Lyer, the perceptual system takes the two-dimensional arrow heads and arrow tails to be depth cues, which the system interprets as indications that the line between the arrow heads is longer than the line between the arrow tails. Thus, it has been frequently argued, size constancy scaling by the perceptual system accounts for the inaccurate representation of the length of the lines.

Palmer points out that these kinds of perceptual illusion cases sometimes compel critics to conclude that perception is largely inaccurate and doubt the veracity of the accuracy view. However, he claims these visual illusion cases are not overly worrisome:

> It is easy to get so carried away by illusions that one starts to think of visual perception as grossly inaccurate and unreliable. This is a mistake. As we said earlier, vision is useful to the extent that it is accurate—or, rather, as accurate as it needs to be. […] The only aspect that is inaccurately perceived is the single illusory property—[in the case of the Müller-Lyer illusion,] the relative lengths of the horizontal lines—and the discrepancy is quite modest. Moreover, illusions such as these are not terribly obvious to everyday life; they occur most frequently in books about perception. *All things considered, then, it would be erroneous to believe that the relatively minor errors introduced by vision overshadow its evolutionary usefulness.* (1999, 8—my emphasis).

---

[15] There is disagreement over the explanation for the Müller-Lyer illusion. For example, Heller, Brackett, Wilson, Yoneyama, Boyer, and Steffen (2002) found that tactile versions of the Müller-Lyer illusion worked for congenitally blind patients, suggesting that the effect is not strictly visual in nature. However, for my purposes it does not actually matter that that size constancy scaling is the correct explanation for the illusion. The size constancy scaling explanation will merely demonstrate how certain examples of perceptual inaccuracy can still be explained by reference to the accuracy view. For more on classic and alternative explanations of the Müller-Lyer illusion see Bermond and Heerden (1996) and Zeman, Obstm Brooks, and Rich (2013).

Thus, he claims that these illusions in fact are relatively minor and not central to the way we experience the world in "everyday life". He goes onto say,

> [P]erceptual errors produced by these illusions may actually be relatively harmless *side effects* of the same processes that produce veridical perception under ordinary circumstances. (ibid, 9—my emphasis) […] [In] most everyday circumstances … normal visual perception is highly veridical. (ibid, 23–24).

The idea here is that the perception is, and functions to be, largely accurate. Though illusions like the Müller-Lyer are cases of inaccurate perceptual representation, the inaccuracy of the illusory perceptual state does not 'count against' the accuracy view because—so the accuracy view proponent can argue—size constancy scaling *typically* generates accurate perceptual representations when we interact with the three-dimensional world. It just so happens that size constancy scaling, which typically generates accurate representations, has an unavoidable *side effect* of producing inaccurate perceptual representations when we look at two-dimensional images with "misleading" depth cues. We can think of 'side effects' like engineering limitations—the organization of a biological system will have been selected for because it (perhaps over all other possible organizations) allows the system to fulfil some function *f*. But there may still be circumstances where this selected-for organization fails to preform *f*. Nature does not always afford perfect engineering solutions and evolution can only 'do the best it can'.

Now armed with the idea of an 'accuracy side effect', consider the commitments of the accuracy view. I take the common version of the accuracy view (which Graham and Palmer seem to endorse) to be what I'll call 'accuracy view monism'. Accuracy view monists think that accuracy is the only and only function of perception. Thus, they are committed to explaining all instances of perceptual inaccuracy in terms of the perceptual system functioning to deliver accurate representation. So, *in*accurate perceptual states will either be accuracy side effects or

malfunctions. This is the version accuracy view that has long been lurking in the shadows of philosophy and cognitive science, which I will herein challenge.

3. Action Guidance

I will explore and defend what I will call the '*action guidance view*', according to which the perceptual system functions to produce perceptual states that guide action in a way that promotes differential reproduction. But how should we understand action guidance? While taking on that question would take us too far afield, I want to situate the action guidance view within the context of action guiding views of representation defended by Anderson and Rosenberg (2008) and Milner and Goodale (1995).

Anderson and Rosenberg defend an action-guiding theory of representation, arguing that *representation itself* is fundamentally in the business of guiding action:

> We hold that *what a representation does* is provide guidance for action. Whatever the details of its instantiation or structure, whatever its physical or informational features (and these are quite various across different representing systems), what makes a given item representational is its role in providing guidance to the cognitive agent for taking actions with respect to the represented object (2008, my emphasis).

Their claim is that representation exists to enable agents to successfully interact with the objects being represented. We represent things, like tigers and philosophy departments, so that we can take actions towards them. Therefore, when determining the structure and function of a representational system, we must look to action and the way the system enables actions (this should sound a lot like the holistic criteria for determining biological functions from section 2!). This aspect of representation can help situation the action guidance view.

But what about *perceptual* representation? Anderson and Rosenberg claim that representational systems are structured in an action guiding way because representation exists to guide action. But we will need to flesh this out a bit more in the perceptual domain to situate

the action guidance view. Anderson and Rosenberg see themselves as building off Milner and Goodale, who describe the way natural selection has shaped vision, stressing that action guidance is central to the visual system (1995, 11, my emphasis):

> Vision in the frog, like vision in other organisms, did not evolve to provide perception of the world in any obvious sense, but rather to provide distal sensory control of the movements that the animal makes in order to survive and reproduce in that world. Natural selection operates at the level of overt behavior; *it cares little about how well an animal 'sees' the world, but a great deal about how well the animal forages for food, avoids predators, finds mates, and moves efficiently from one part of the environment to another*. To understand how the visuomotor systems controlling these behaviors are organized, *it is necessary to study both the selectivity of their sensory inputs and the characteristics of the different motor outputs* they control.

Note that Milner and Goodale make a point (similar to Burge's) that natural selection operates at the level of overt behavior (because overt behavior influences differential reproduction) and is not specifically concerned with the phenomenology of internal representational states—or rather, is not concerned with the phenomenology of internal states for the sake of the phenomenology of internal states. They emphasize that, at the level of overt behavior, what matters for the visual system is enabling actions that promote an organism's differential reproduction (foraging for food, avoiding predators, finding mates, moving from one part of the environment to another).

Where does that leave the action guidance view with regards to the accuracy view? Let's review the ground covered so far. Biological systems (perception included) are organized via evolution by natural selection and evolution by nature selection operates at the level of *overt, action-guiding behavior*. But recall from our discussion of Burge and Graham in section 2 that it might well be that the perceptual system is organized via evolution by natural selection to produce *accurate* internal representational states because it is contingently true that accurate perceptual states guide action in a way that enables differential reproduction. This would essentially mean that the accuracy view and the action guidance view would collapse into one

another, which is what monistic accuracy view defenders like Graham and Palmer seem to assume.

What sort of evidence would be required to suggest that the action guidance does *not* collapse into accuracy view? There are a couple things worth flagging here. First, if we could find a class of perceptual states that are inaccurate, but which could not be claimed as accuracy side effects or cases of perceptual malfunction, then we could assume that there is a function of perception that is *not* accurate representation (if we found perceptual states are inaccurate, not perceptual malfunctions, and not accuracy side effects, then the states must be either fulfilling a non-accuracy function of the perceptual system or be the side effect of non-accuracy function of the perceptual system). Second, it we could find empirical evidence which suggest that there is a non-accuracy function of perception and we judged that action guidance is the function the perceptual system is in these instances fulfilling, then we could infer to the best explanation that: (1) action guidance is the (or, *is at least one of the*) function(s) of perception, (2) the action guidance view is not nested in the accuracy view, and (3) the monistic accuracy view is wrong. I will argue that the attention effects in the next section provide examples for one such class of perceptual states. For clarity, I've sketch out the ground we have covered in the first three sections:[16]

> P1. A perceptual state *s* accurately represents the world, if and only if *s* represents the world as being χ and the world is in fact that way (χ). (Definition of perceptual accuracy)

[16] An anonymous reviewer has questioned why I formally lay out my premises since there's no complicated logic going on. This argument is long and complicated so I think seeing premises laid out like this can help readers keep track of what's going on. If it doesn't help you, feel free to skip over these sections going forward.

P2. Perceptual states are either accurate or inaccurate. (Assumption)

P3. Producing accurate perceptual states is the one and only function of the perceptual system. (The monistic version of the accuracy view)

P4. If producing accurate perceptual states is the one and only function of the perceptual system, then *inaccurate* perceptual states are either malfunctions or accuracy side effects (see 2.1 for discussion of side effects).

C1. Therefore, according to the accuracy view, inaccurate perceptual states are either malfunctions or accuracy side effects of the perceptual system.

C2. If we discovered a class of perceptual states that are inaccurate but are neither malfunctions nor accuracy side effects, then the accuracy view is wrong and there must be another function of perception.

4. Action Guidance as Perceptual Function: Consulting the Empirical Data

4.1. Endogenous and exogenous attention

Can attention affect the way objects appear? Both Hermann Helmholz (1866) and William James (1890/1983) thought that attention altered and intensified the appearance of attended-to objects. More recent research over the last 10–15 years has suggested that indeed both endogenous attention (voluntary attention) and exogenous attention (involuntary, transient attention, usually triggered by a change in the periphery that is reflexively attended to) can

affect the appearance of objects. The studies reveal that attention systematically distorts the appearance of objects across a variety of dimensions.

Exogenous attention affects perception of contrast (Carrasco, Ling, & Read 2004; Carrasco, Fuller, & Ling 2008; Ling, & Carrasco 2007; Fuller, Rodriguez, & Carrasco 2008; Störmer, McDonald, & Hillyard 2009), color saturation (Fuller & Carrasco 2006), motion coherence (Liu, Fuller, & Carrasco 2006), object size (Anto-Erxleben, Henrich, & Treue 2007), speed (Turatto, Vescovi, & Valsecchi 2007; Fuller, Park, Carrasco 2009), gap size and spatial frequency (Gobell & Carrasco 2005), and flicker rate (Montagna, & Carrasco 2006). *Endogenous attention* affects perception of contrast (Liu, Abrams, & Carrasco 2009), brightness (Tse 2005), and spatial frequency (Abrams, Barbot, & Carrasco 2010). Rahnev and Denison (2016) summarize these findings, saying that "directing spatial attention to a stimulus can make it appear higher contrast… larger… faster… brighter… and higher spatial frequency than it would otherwise" (20).

For the rest of the paper, I will discuss two studies in particular—Carrasco, et al. (2004) and Liu et al. (2009)—which tested the effects of endogenous and exogenous attention on perception of contrast and will argue that these effects (which I will collectively refer to as the 'attention effects') are best accounted for by the action guidance view. While my discussion will focus on the way exogenous and endogenous attention specifically effects perception of contrast, it's worth flagging that I do think similar kinds of action guiding explanations could be used to account for some of the other attention effects cited above.[17]

[17] Jake Beck has objected to my action-guiding interpretation of the contrast boost, arguing that while perhaps the action guiding explanation might seem reasonable in the contrast case, not all of the purported attention effects lend themselves as well to an action guiding explanation, citing in particular the effect of both endogenous and exogenous attention on spatial frequency. However, I do not take myself to being committed here to arguing that *every* attention effect was selected for because it guides action in a survival promoting way. Nonetheless, it seems plausible to think that some of the attention effects may have been selected for and some might be 'side effects' of the other, selected for effects. I would argue that the object size and color saturation particularly lend themselves to an action guiding explanation.

4.2. The data

Carrasco et al. (2004) tested effects of *exogenous* attention on perceived contrast.[18] I have chosen to discuss this particular study in some depth because it is frequently discussed in the empirical literature (and in philosophy—Block, 2010) and because readers will be able to appreciate the perceived contrast differences of the stimuli (the other kinds of cited attention affects are less easy to appreciate on paper).

Experimental design: Participants were told to report the orientation of the Gabor patch (the tilted circular grids) with the higher contrast. "To preclude response bias… [t]he experimental design emphasized to observers the orientation judgment, when in fact we were interested in their contrast judgments", Carrasco et al. (2004) stated. Therefore, participants were not aware experimenters were testing their perception of contrast. Participants shown a gray screen for 500ms that had a fixation point (a black dot) in the center and were instructed to focus on the center of the screen for the duration of the experiment. A cue then briefly appeared on the left, right, ('peripheral' cues) or center of the screen ('neutral' cue), automatically drawing the participants' exogenous attention to the area of the screen where the cue was shown. Note that participants were told before the experiment began that the location of the peripheral cue would not be related to the location or orientation of the higher contrast Gabor patch. Following a 53ms interstimulus interval (ISI), participants were then shown two Gabor patches that appeared for 40ms on the left and right of the fixation point. They were then asked, "is the stimulus that looks higher in contrast tilted to the right or left?" and had to indicate the orientation of the Gabor patch with the highest contrast via a keypress. An

[18] Though discussion of this alternative interpretation would take us too far afield, it is worth noting that Beck and Schneider (2017) deny the empirical premise that attention is affecting perception of contrast at all in these experiments and attribute the effect instead to salience. However, for the sake of this paper, I follow the majority interpretation and assume the effect that attention is altering contrast and not salience.

inferred camera was used to detect participants' eye movements, ensuring that they did not look away from the fixation point.

Carrasco et al. (2004) found that if two identical Gabor patches were shown, participants saw the cued patch as being higher in contrast than the uncued patch. They also found that if the cued patch was lower in contrast than the uncued patch, then attending to the cued patch caused participants to perceive the patches as being of equal contrast. By showing pairings of Gabor patches, the experimenters were able to determine the effect of exogenous attention at various levels of contrast on perceptual phenomenology (i.e. they were able to map out what the contrast difference between cued and uncued patches would have to be for the participants to perceive the patches as equal in contrast). For example, Carrasco et al. found that a cued 22% and an uncued 28% Gabor patch looked identical to participants, meaning that exogenous attention was worth 6% contrast. As Block notes of these experiments, in addition to a contrast boost of the attended-to patch, "this effect no doubt involves decreased apparent contrast of the less attended to patch". Thus, for the sake of this paper I will assume that the effect involves some combination of a contrast boost of the attended-to stimuli and a contrast reduction of the unattended-to stimuli.[19] Though 6% might not sound like a lot, a quick glance at the Gabor patches will reveal that a 6% difference is fairly significant. An effect of this magnitude could indeed be the difference between being able to identify and respond to a briefly presented stimuli and failing to identify it altogether (much more about contrast boost and object identification in section 5).

[19] The Carrasco et al. (2004) experimental design does not allow enable one to distinguish the percentage of the effect that is caused by a perceived contrast boost to the attended-to stimuli and the percentage of the effect that is caused by a perceived contrast reduction to the unattended-to stimuli. However, while I make Block's assumption and refer to the effect as some combination of a boost and reduction, my argument should equally apply if the effect was wholly the result of a contrast boost or a contrast suppression. Therefore, insert your preferred interpretation of the effect where appropriate.

But what about *endogenous* attention? Does voluntarily attending to a stimulus make it appear higher contrast? Liu, Abrams, and Carrasco (2009) slightly modified the Carrasco et al. (2004) experimental design and found that voluntary endogenous attention boosts perceived contrast by similar magnitudes. "Does voluntarily attending to an object alter its appearance? Our results indicate that the answer to this age-old question is 'yes.'", Liu et al. (2009) write, "the increase in apparent contrast observed with voluntary attention parallels results found with involuntary [exogenous] attention." Therefore, both studies indicated that there a similar overall effect of attention on perceptual phenomenology. I will expound on the significance of endogenous and exogenous attention exhibiting the same effect of perceptual phenomenology in 4.6.

Hereafter, 'attention effects' will refer to the endogenous and exogenous attention causing agents to perceptually experience attended-to stimuli as higher in contrast and unattended-to stimuli as lower in contrast.

4.3 Interpreting the data: Are the attention effects genuinely perceptual?

For the sake of this paper I am *assuming* that the effects are genuinely perceptual and are *not* caused by some form of post-perceptual cognitive judgment. Ned Block argues for this interpretation of the attention data, asserting that "it has been settled beyond any reasonable doubt that the effect is a genuine perceptual effect rather than any kind of cognitive effect" (2010, 37). Block cites a variety of compelling reasons to think this, all of which I will not review here. One point he makes is particularly relevant to our discussion, however. In reference to the Carrasco et al. (2004) findings, Block emphasizes the importance of the temporal duration of the effect. Exogenous attention peaks at 100ms and decays soon after, which means that if the effect were connected to exogenous attention, we would expect that it would decay after 100ms. And indeed Carrasco et al. (2004) found that when the cue and

the stimulus are presented 53ms apart, the effect manifests. However, when the cue and the stimulus are presented 500ms apart, the effect disappears. Thus, because the effect disappears during the period of exogenous attention decay, we have good reason to think that exogenous attention is responsible for the effect rather than a cognitive bias. If, on the other hand, the effect was being driven by a cognitive bias, we would expect that it would take more time to manifest and would not quickly disappear between 53ms and 500ms. So, because the effect manifests when the cue and stimulus are 53ms apart and disappears when the they are 500ms apart, I will assume the effects are genuinely perceptual.

4.4 Interpreting the data: What about accuracy?

So, the attention effects are perceptual. But are they examples of inaccurate representation? Recall that accuracy involves a perceptual state accurately representing the world as being χ and the world is in fact that way (χ). Perceptual accuracy is thus pretty straightforward—a perceptual representation is accurate if it represents the world as being the way it in fact is. If a perceptual state has the content of representing the Gabor patch as having a contrast of 16% but the Gabor patch in fact has a contrast of 22%, then the perceptual state is representing the world as being a way that it is not, therefore the perceptual state is inaccurately representing the world.

I think there is intuitive force behind this interpretation of the attention effects being both representational and inaccurate. For one, empirical researchers in the field write about the effects in decidedly representational terms (Liu et al.: "our results showed that voluntary attention increased perceived contrast"; Carrasco et al.: "transient attention increases apparent contrast for a wide range of stimulus contrasts"), suggesting they take the effects to be representational. Furthermore, given the definition of accurate perceptual representation, it seems most natural to interpret the effects as genuine examples of inaccurate perceptual

representation. However, it is worth flagging that there are those that deny that effects are representational at all. Block appeals to the idea of 'mental paint', arguing that the phenomenological changes brought about by attention cannot be explained by changes in representational content of the perceptual states (2010). I would argue that a 22% patch appearing 28% does absolutely involve a change in representational content—attention causes the Gabor patch to be represented as 28% rather than 22%. But, for the sake of discussion in this paper, I will not discuss the alternative non-representational views and assume that the effects are indeed genuine examples of inaccurate representation.

4.5 Putting it together

Recall C1 and C2:

> C1. Therefore, according to the accuracy view, inaccurate perceptual states are either malfunctions or accuracy side effects of the perceptual system.
>
> C2. If we discovered a class of perceptual states that are inaccurate but are neither malfunctions nor accuracy side effects, then the accuracy view is incorrect and there must be another function of perception.

So, the exogenous and endogenous attention effects are (1) perceptual and (2) involve inaccurate perceptual representation. What does that mean for our discussion of the function of perception? Consider the accuracy view. If the accuracy view monist accepts that attention effects are perceptual and inaccurate, then by producing perceptual states that inaccurately represent the contrast of the Gabor patches, the perceptual system is failing to fulfil its one and only function. To preserve the accuracy view, the accuracy view monist must claim that

the attention effects are either accuracy side effects or malfunctions. I will argue that both claims are implausible.

4.6 Attention effects as side effects

To save the accuracy view, the accuracy view monist could claim that the attention effects are accuracy side effects. But this explanation runs into difficulty.

Recall that endogenous attention and exogenous attention effect perceived contrast in largely the same way (i.e. to the same degree at different levels of contrast). Liu et al. (2009) note the oddity of this given the fact that endogenous and exogenous attention have "different time courses and control processes, as well as different effects on perceptual performance" (360). They go on to say that "it is not obvious why two such different forms of attention would have similar phenomenological consequences" and they do not provide an explanation for the striking similarity between the endogenous and the exogenous contrast boosts. However, I contend that because (1) the endogenous and exogenous attention effects have almost exactly the same effect on perceptual phenomenology and (2) many theories of attention suggest that endogenous and exogenous attention are governed by independent systems, there is good reason to think that the endogenous and exogenous contrast boosts are not accuracy side effects. This claim will require some background and further unpacking.

While it was previously thought that endogenous and exogenous attention were features of a single unified attention system, contemporary research has put pressure on this theory. For a thoroughgoing review of the literature from psychology and neuroscience on the evidence for the existence of two separate attention systems see Chica, Bartolomeo, and Lupiáñez (2013). Summarizing their review of the literature, they assert that "accumulating behavioral evidence indicates that endogenous and exogenous attention differ not only in

quantitative aspects (such as the magnitude of the attentional effects or their time course), but also on their qualitative effects on information processing" (119). Exogenous attention affects the earliest stages of visual processing, enhancing the appearance of the stimuli and processing of object-related information, whereas endogenous attention involves processing of spatial location information and reduction of irrelevant background noise. Therefore, they conclude, "the accumulation of behavioral dissociations of the effects of endogenous and exogenous attention gives strong support to the hypothesis that *endogenous and exogenous attention consist of two independent attentional systems, with well differentiated functional characteristics*" (Chica et al. 2013, 120, my emphasis). Relatedly Mayer, Dorflinger, Rao, and Seidenberg (2004) found that endogenous and exogenous attention did not even have common neuronal substrates, as was once thought. They found that *exogenous attention* activated the bilateral temporal juncture, bilateral superior temporal gyrus, right middle temporal gyrus, right frontal eye field, and left superior temporal gyrus, and *endogenous attention* activated only the left superior temporal gyrus (534).[20] So there is reason to think that not only are the two types of attention functionally distinct but they are facilitated by entirely different brain regions.

Why might the distinctness of the attention systems matter to the question of the attention effects being accuracy side effects? If the systems are functionally distinct and facilitated by different brain regions (that, according to the accuracy view monist, must function to produce accurate perceptual representations), then it seems odd that both systems would happen to have the *same* accuracy side effect. For the accuracy view monist to maintain that the attention effects are accuracy side effects, they would need two different stories (one for the endogenous attention system and one for the exogenous attention system) as to why

[20] In this study, they used event-related FMRI (ER-FMRI) to examine the brain regions associated with endogenous and exogenous attention (Mayer et al. 2004).

the contrast boost is an accuracy side effect of each system. They would then need to chock the similarity between the two side effects up to coincidence (if we accept that there is not a functional or physical similarity between the two attention systems that can be appealed to). However, because the two neurologically and functionally distinct systems affect perceptual phenomenology in the same way, it might, perhaps, be reasonable to think that the attention effects are not merely side effects that both the exogenous attention system and endogenous attention system coincidentally share.

4.7 Attention effects as malfunctions

Accuracy monists could still come back and say that perhaps endogenous and exogenous attention effects are malfunctions of the perceptual system. However, if this is right and the attention effects are just perceptual malfunctions, then it will be the case that our perceptual system is constantly in the process of systematically malfunctioning. I think this claim should strike us as fairly improbable. Attention effects are not like an esoteric perceptual illusion from a vision science textbook. We are constantly in the process endogenously and exogenously attending to objects in the world. If a theory of perceptual function entails that our perceptual system is *constantly* in the process of malfunctioning, then it is reasonable to think that we should be highly suspicious of the theory. For these reasons, diagnosing the attention effects as perceptual malfunctions seems unsatisfactory at best.

So, the accuracy monist must claim that the attention effects are either accuracy side effects or malfunctions. Both of these explanations are empirically and theoretically fraught. How then should we account for the attention effects? I think this is where the action guidance view can help. Building off of the argument formalization in 4.5, which ended with C1, this is the further ground we've covered:

C1. Therefore, according to the accuracy view, inaccurate perceptual states are either malfunctions or accuracy side effects of the perceptual system.

C2. If we discovered a class of perceptual states that are inaccurate but are neither malfunctions nor accuracy side effects, then the accuracy view is incorrect and there must be another function of perception.

P5. Endogenous and exogenous attention inaccurately distort perceptual states in a variety of ways—as demonstrated by the 'attention effects'. (see 4.1 for discussion of the Carrasco et al. 2004 and Liu, et al. 2009 data)

P6. The attention effects are neither accuracy side effects nor malfunctions. (see 4.6 and 4.7 for the discussion of malfunctions and side effects)

C3. Therefore, there must be another function of perception that the attention effects can be explained in terms of.

5. Attention and the Action Guidance View

I will argue that the action guidance view is in a better position to account for the exogenous and endogenous attention effects than the accuracy view. Specifically, by producing inaccurate perceptual states (an effect of endogenous and exogenous attention), the perceptual system is *fulfilling its function to produce action guiding perceptual states that promote differential reproduction*. But how would the attention effects guide action in this way? We would indeed need a compelling explanation for how the contrast boost/reduction could promote differential reproduction to

make us seriously entertain the possibility that the effects were selected for. In what follows, I have tried to put forth one such explanation. And though all that is said here far from settles the matter, I hope these theoretical and empirical considerations compel readers to take seriously the possibility that the attention effects (1) were selected for and (2) are best explained within the framework of the action guidance view.

But how do the attention effects relate the action guidance? To answer this question, we must consider for a moment how a contrast increase/decrease associated with the attention effects would impact a perceiver's experience of a visual scene. For one, *increases* in contrast make lines, colors, and contours *more* intense and discriminable and *decreases* in contrast make objects (and features of objects) *less* clear and noticeable. This, coupled with the contention that attended-to objects appear higher in contrast and unattended-to objects appear lower in contrast, means that the attended-to objects will particularly stand out to the perceiver against the backdrop of unattended-to objects. More generally we can understand attention as tracking potentially significant aspects of the visual scene—exogenous attention automatically tracks relevant changes in the scene (for example, a loud sudden noise to the perceiver's left would draw involuntary attention to that direction) and endogenous attention tracks aspects of the scene on which the perceiver places conscious importance (for example, I consciously attend to the traffic light before it changes). Thus, we can understand the attention effects as making potentially significant aspects of a perceptual scene clearer and more noticeable relative to more insignificant aspects. I will argue this promotes action guiding object identification and ultimately differential reproduction.

When an attended-to object is perceptually represented as being higher in contrast, the perceiver is able to more quickly and effectively identify the object. Think about trying to locate a sunflower in a field of wildflowers. The task would take less time if the sunflower was

higher in contrast and the other flowers were lower in contrast. And when perceivers are able to more quickly and effectively identify objects, they are able to act in ways that better promote their differential reproduction. Is the object food, a predator, an injured compatriot? All of these objects will be associated with a different set of affordances. The more quickly a perceiver is able to determine if an object is a rabbit or a tiger, the better chance the perceiver will have to act in ways that promote their continued survival and reproduction. Thus, inaccurately representing the tiger to be higher in contrast (relative to the less important unattended-to aspects of the scene, which appear lower in contrast) allows the perceiver to more quickly pick out and identify the tiger and act according. A couple things about this purported advantageous contrast boost are worth flagging.

For one, obviously a contrast boost will not *always* be needed to successfully identify an object.[21] One does not generally require a contrast boost to be able to quickly identify a giant red barn (unless you are in barn façade county, in which case you would need more than that!); an accurate perceptual state would easily suffice in the task of identifying a giant red barn. Note that in the attention experiments discussed, the stimuli were only very briefly presented to the participants. It would make sense in cases like these, where perceivers momentarily saw the attended-to object, that a slight boost in contrast (making lines, colors, and contours more intense and discriminable) would help them identify the object. Granted, if I am able to look at the object for a long time, something like a 6% contrast difference would not probably seem as important. But if I only see an object for 40ms, *anything* that might help me identify the object would be beneficial. And in fact, in our day-to-day lives we often only see objects very briefly and are forced to make quick judgments about their nature and identity.

[21] Thanks to Eric Mandelbaum for pressing me on this point.

Second, it's certainly true that not every kind of contrast boost would be advantageous. A 70% contrast boost would probably become a perceptual impediment and would not enable identification of relevant features of the attended-to object. But we can imagine that a 6% contrast difference might enable object identification and thus promote differential reproduction. It is reasonable to think that a 6% contrast boost is significant enough to *aid* in object identification but not so significant that it *impedes* visual processing of the scene.

Where does this leave us with regards to the action guidance view? Given the most compelling interpretation of the empirical data and the considerable benefit conferred by the contrast boost/reduction, I submit that we have reason to think that the attention effects are an example of the perceptual system inaccurately representing the world *because* this type of inaccurate representation guides action in a way that promotes differential reproduction. Thus, assuming some version of the teleological account of biological function, the perceptual system would be functioning to guide action. The action guidance view is—at least with respect to attention—venerated.

Furthermore, if my explanation that the attention effects were selected for because perceptual states with the contrast boost/reduction tend to guide action in a way that promotes differential reproduction is correct, then this might explain *why* the functionally and physically distinct exogenous and endogenous attention systems effect perceived contrast by similar magnitudes. This is the contrast percentage boost/reduction that accentuates the pertinent features of directly and indirectly attended-to objects, enabling successful identification and relevant action.

But I anticipate the following objection at this juncture: 'Aha, you are talking about object identification! Object identification involves aiming to make accurate judgments about the world, so your view really is just another version of the accuracy view.' I would respond

by pointing out that this sort of objection conflates two importantly different kinds of accuracy—*perceptual accuracy* and *cognitive accuracy*.

As we've discussed, perceptual accuracy is pretty straightforward. Does a perceptual state represent the world as it in fact is? If it does, then the perceptual state is accurate. If it does not, then the perceptual state is inaccurate. However, the present objection conflates perceptual accuracy with accurate object identification, which is a kind of *cognitive* accuracy. Object identification involves making a (conscious or non-conscious) *cognitive judgment* about the representational contents of a perception state. If I identify the furry object on my couch—determining that 'the object on the couch is a cat'—then I am making a cognitive judgment.[22] Of course, perceptual states will very often influence cognitive judgments. But object identification involves making a cognitive judgment *about* the identity of some perceptually represented object or stimuli. Hence, *accurate object identification* involves judging that some perceived object is an *X* and it being that case that the object is an *X*.

Thus, I am absolutely asserting that perceptual inaccuracies like the attention effects can promote accurate object identifications that guide action in a way that promotes differential reproduction. In fact, I have here suggested that the (inaccurate and representational) contrast boost/reduction promotes accurate object identification better than accurate representation would, giving us compelling reasons to think the attention effects were selected for. However, it is important for my argument the perceptual state in question is *representationally inaccurate* as a result an endogenous or exogenous attention effect.

---

[22] For more on object recognition see Hope (2009), Smith (2003), and Farah (1992).

6. Conclusion

It is clear that attention is absolutely fundamental to how we interact with, perceive, and understand the world. And research over the last couple decades has helped us establish that attention inaccurately distorts our perceptual states in systematic ways, which I argue increases the chances of differential reproduction. Thus, reflecting on the relationship between perception and attention, we should want our account of perceptual function to sensibly accommodate the numerous and well documented effects of exogenous and endogenous attention on perceptual phenomenology. I have suggested that the action guidance view is in a strong position to do just this. Let's turn back to where we left off:

> C2. If we discovered a class of perceptual states that are inaccurate but are neither malfunctions nor accuracy side effects, then the accuracy view is wrong and there must be another function of perception.
>
> P5. Endogenous and exogenous attention inaccurately distort perceptual states in a variety of ways—as demonstrated by the 'attention effects'. (see 4.1 for discussion of the Carrasco et al. 2004 and Liu, et al. 2009 data)
>
> P6. The attention effects are neither accuracy side effects nor malfunctions. (see 4.3 for the discussion of malfunctions and side effects)
>
> C3. Therefore, there must be another function of perception that the attention effects can be explained in terms of.

In light of the considerations brought up in section 5, we can now add P7, P8, and C4:

P7. The available empirical data indicates attention effects aid in quick and successful object identification in way that confers action guiding differential reproduction benefit.

P8. If the available empirical data indicates attention effects aid in quick and successful object identification in way that confers action guiding differential reproduction benefit, then we can infer to the best explanation that the attention effects can be best explained in terms of the perceptual system functioning to produce action guiding perceptual states. (see argument for this in section 5)

C4. Therefore, the perceptual system functions to produce action guiding perceptual states.

Of course, future research about the relationship between object identification and stimuli contrast will hopefully shed more light on the nature of the attention effects. But, as the data stands, the attention effects lend compelling support for the action guidance view. Moreover, given that the accuracy view monism is often assumed without defense (or it is assumed that the accuracy view and the action guidance view are essentially one in the same), one can hope that the action-centered views of biological function (like the action guidance view) will come be viewed more favorably, given their ability to accommodate data for which accuracy views have difficulty accounting.

BIBLIOGRAPHY

Abrams, J., Barbot, A., & Carrasco, M. (2010). Voluntary attention increases perceived spatial frequency. *Attention, Perception, & Psychophysics*, 72(6), 1510-1521.

Aiden, H. & McCarthy, A. (2014). Current attitudes towards disabled people. London: Scope.

Allen, C. (1992). Mental content. *The British Journal for the Philosophy of Science*, 43(4), 537-553.

Allport, G. (1954). *The Nature of Prejudice*. Oxford, England: Addison-Wesley.

Alter, A. & Oppenheimer, D. (2009). Uniting the tribes of fluency to form a metacognitive nation. *Personality and Social Psychology Review*, 13(3), 219-235.

Alter, A., Oppenheimer, D. , Epley, D., Epley, N., Eyre, R. (2007). Overcoming intuition: Metacognitive difficulty activates analytic reasoning. *Journal of Experimental Psychology: General*, 136(4), 569-576.

Anton-Erxleben, K., Henrich, C., & Treue, S. (2007). Attention changes perceived size of moving visual patterns. *Journal of Vision*, 7(11), 5, 1-9.

Aslin, R., Saffran, J., & Newport, E. (1998). Computation of conditional probability statistics by 8-month-old infants. *Psychological Science*, 9(4), 321-324.

Baker, C., Olson, C., & Behrmann, M. (2004). Role of attention and perceptual grouping in visual statistical learning. *Psychological Science*, 15(7), 460-466.

Bauer, P. (1993). Memory for gender-consistent and gender-inconsistent event sequences by twenty-five-month-old children. *Child Development*, 64(1), 285-297.

Beck, J., Schneider, K. (2017). Attention and mental primer. *Mind & Language*, 32(4), 463-494.

Bermond, B., Heerden, J. (1996). The Müller-Lyer illusion explained and its theoretical importance reconsidered. *Biology and Philosophy*, 11(3), 321-338.

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.

Block, N. (2010). Attention and mental paint. *Philosophical Issues*, 20(1), 23-63.

Block, N. (2014). Seeing-as in the light of vision science. *Philosophy and Phenomenological Research*, 89(3), 560-572.

Boghossian, P. (1995.) Content. In *A Companion to Metaphysics*, eds. J. Kim & E. Sosa. Oxford: Blackwell.

Burge, T. (2010). *Origins of Objectivity*. Oxford: Clarendon Press.

Carey, S. (2009). *The Origin of Concepts*. Oxford: Oxford University Press.

Carrasco, M., Fuller, S., & Ling, S. (2008). Transient attention does increase perceived contrast of suprathreshold stimuli: A reply to Prinzmetal, Long and Leonhardt (2008). *Perception & Psychophysics*, 70(7), 1151-1164.

Carrasco, M., Ling, S., & Read, S. (2004). Attention alters appearance. *Nature Neuroscience*, 7(3), 308-313.

Chica, A., Bartolomeo, P., & Lupiáñez, J. (2013). Two cognitive and neural systems for endogenous and exogenous spatial attention. *Behavioral Brain Research* 237, 107-123.

Cohen, J. (2015). Perceptual representation, veridicality, and the interference theory of perception. *Psychonomic Bulletin & Review*, 22(6), 1512-1518.

Coren, S. (1986). An efferent component in the visual perception of direction and extent. *Psychological Review*, 93(4), 391-410.

Courcille A., Daw, N., & Touretzsky, D. (2006). Bayesian theories of condition in a changing world. *Trends in Cognitive Science*, 10(7), 294-300.

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

Crump, M., McDonnell, J., & Gureckis, T. (2013). Evaluating Amazon's Mechanical Turk as a tool for experimental behavioral research. *PLoS One*, 8(3), e57410.

De Houwer, J., Teige-Mocigemba, S., & Moors A. (2009). Implicit measures: A normative analysis and review. *Psychological Bulletin*, 135(3), 347-368.

Dennett, D. (1969). *Content and Consciousness*. London: Routledge.

Dennett, D. (1983). Styles of mental representation. *Proceedings of the Aristotelian Society*, 83(1982–1983), 213-226.

Dovidio, J., Hewstone, M., Glick P., & Esses V. (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. Fragmentation and information access. ms

Evans, J. (2006). The heuristic-analytic theory of reasoning: Extension and evaluation. *Psychonomic Bulletin & Review*, 13(3), 378-395.

Evans, J., Barston, J., & Pollard, P. (1983). On the conflict between logic and belief in syllogistic reasoning. *Memory & Cognition*, 11(3), 295-306.

Feldman, J. (2015). Bayesian models of perceptual organization. In *The Oxford Handbook of Perceptual Organization*, ed. J. Wagemans. Oxford: Oxford University Press.

Field, H. (1978). Mental representation. *Erkenntnis*, 13(1), 9-61.

Firestone, C., & Scholl, B. (2016). There are no top-down effects. *Behavioral and Brain Sciences*, 39, 1-72.

Fiske, S. (1998). Stereotyping, prejudice, and discrimination. In *The Handbook of Social Psychology*, eds. D. Gilbert, S. Fiske, & G. Lindzey. New York, NY: McGraw-Hill, 357-411.

Fodor, J. A. (1985). Fodor's guide to mental representation: The intelligent auntie's vade-mecum. *Mind*, 94(373), 76-100.

Fuller S., & Carrasco M. (2006). Exogenous attention and color perception: Performance and appearance of saturation and hue. *Vision Research*, 46(23), 4032-4047.

Fuller S., Park Y., & Carrasco M. (2009.) Cue contrast modulates the effects of exogenous attention on appearance. *Vision Research*, 49(14), 1825-1837.

Fuller, S., Rodriguez, R. Z., & Carrasco, M. (2008.) Apparent contrast differs across the vertical meridian: Visual and attentional factors. *Journal of Vision*, 8(1), 1-16.

García-Carpintero, M. (1994). The supervenience of mental content. *Proceedings of the Aristotelian Society*, 94, 117-135.

Geisler, W. S., & Diehl, R. L. (2003.) A Bayesian approach to the evolution of perceptual and cognitive systems. *Cognitive Science*, 27(3), 379-402.

Gilbert, D. & Hixon, J. (1991). The trouble of thinking: Activation and application of stereotypic beliefs. *Journal of Personality and Social Psychology*, 60(4), 509-517.

Gobell, J., & Carrasco, M. (2005). Attention alters the appearance of spatial frequency and gap size. *Association of Psychological Science*, 16(8), 644-651.

Godfrey-Smith, P. (1984.) A modern history theory of functions. *Noûs*, 28(3), 344-62.

Graham, P. J. (2010.) Testimonial entitlement and the function of comprehension. In *Social Epistemology*, eds. A. Haddock, A. Millar, & D. Pritchard. Oxford: Oxford University Press.

Graham, P. J. (2012.) Epistemic entitlement. *Noûs*, 46(3), 449-482.

Graham, P. J. (2014a). Functions, warrant, history. In *Naturalizing Epistemic Virtue*, ed. A. Fairweather. New York, NY: Cambridge University Press.

Graham, P. J. (2014b). The function of perception. In *Virtue Epistemology Naturalized: Bridges Between Virtue Epistemology and Philosophy of Science*, ed. A. Fairweather. Dordrecht: Springer.

Griffiths, P. (1993). Functional analysis and proper functions. *British Journal for the Philosophy of Science*, 44(3), 409-422.

Helmholtz, H. (1866). *Treatise on Physiological Optics*. Trans., J. P. C. Southall. Rochester, NY: Optic Society of America.

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

Howarth, S., Handley, S., & Walsh, C. (2016). The logical-bias effect: The role of effortful processing in the resolution of belief-logic conflict. *Memory & Cognition*, 44(2), 330-349.

Hunt, R. & Aslin, R. (2001). Statistical learning in a serial reaction time task: Access to separable statistical cues by individual learners. *Journal of Experimental Psychology: General*, 130(4), 658-680.

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.

James, W. (1983). *The Principles of Psychology*. Original work published 1890. Cambridge, MA: Harvard University Press.

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

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

Ling, S., & Carrasco, M. (2007). Transient covert attention does alter appearance: A reply to Schneider (2006). *Perception & Psychophysics*, 69(6), 1051-1058.

Lippman, W. (1922). *Public Opinion*. New York, NY: Harcourt, Brace and Company.

Liu T., Abrams, J., & Carrasco, M. (2009). Voluntary attention enhances contrast appearance. *Association of Psychological Science*, 20(3), 354-362.

Liu, T., Fuller, S., & Carrasco, M. (2006). Attention alters the appearance of motion coherence. *Psyhonomic Bulletin & Review*, 13(6), 1091-1096.

Mamassian, P., Landy, M., & Laurence, M. (2002). Bayesian modeling of visual perception. In *Probabilistic Models of the Brain: Perception and Neural Function*, eds. R. Rao, B. Olshausen, & M. Lewicki. Cambridge, MA: MIT Press, 13-36.

Mandelbaum, E. (2015). Attitude, inference, association: On the propositional structure of implicit bias. *Noûs*, 50(3), 629–658.

Mandelbaum, E. (2018). Seeing and conceptualizing: Modularity and the shallow contents of perception. *Philosophy and Phenomenological Research*, 97(2), 267-283.

Mayer, A., Dorflinger J., & Seidenberg, M. (2004). Neural networks underlying endogenous and exogenous visual-spatial orienting. *Neuroimage*, 23(2), 534-541.

Mellor, D. H. (1988). Crane's waterfall illusion. *Analysis*, 48(June), 147-150.

Miller, C., Trautner, H., & Ruble, D. (2006). The role of gender stereotypes in children's preferences and behavior. In *Child Psychology: A Handbook of Contemporary Issues*, eds. L. Balter & C. Tamis-LeMonda. New York City, NY: Psychology Press, 293-323.

Millikan, R. (1989). In defense of proper function. *Philosophy of Science*, 56(2), 288-302.

Millikan, R. (2002). Biofunctions: Two paradigms. In *Functions: New Readings in the Philosophy of Psychology and Biology*, eds. R. Cummins, A. Ariew, & M. Perlman. Oxford: Oxford University Press.

Misyak, J. & Christiansen M. (2012). Statistical learning and language: An individual differences study. *Language Learning*, 62(1), 302-331.

Montagna, B., & Carrasco, M. (2006). Transient covert attention and the perceived rate of flicker. *Journal of Vision*, 6(9), 955-965.

Moskowitz, G., Li, P., & Kirk, E. (2004). The Implicit Volition Model: On the preconscious regulation of temporarily adopted goals. In *Advances in Experimental Social Psychology*, Vol. 36, ed. M. Zanna. San Diego, CA: Elsevier Academic Press, 317-413.

Munton, J. (forthcoming). Perceptual skill and social structure. *Philosophy and Phenomenological Research*.

Müller, F. & Rothermund, K. (2014). What does it take to activate stereotypes? Simple primes don't seem enough: A replication of stereotype activation (Banaji & Hardin, 1996; Blair & Banaji, 1996). *Social Psychology*, 45(3), 187-193.

Neander, K. (1991). Functions as selected effects. *Philosophy of Science*, 58(2), 168-184.

Noë, A. (2005). Real presence. *Philosophical Topics*, 33(1), 235-264.

Noë, A. (2009). *Out of our Heads*. New York, NY: Hill and Wang.

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

Orlandi, N. (2014). *The Innocent Eye: Why Vision is not a Cognitive Process*. Oxford: Oxford University Press.

O'Regan, J. & Noë, A. (2001). A sensorimotor account of vision and visual consciousness. *Behavioral and Brain Sciences*, 24(5), 939-973.

O'Regan, J. (2011). *Why Red Doesn't Sound Like a Bell: Explaining the Feel of Consciousness*. Oxford: Oxford University Press.

Palmer, S. (1999). *Vision Science: Photons to Phenomenology*. Cambridge, MA: MIT Press.

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, NY: Oxford University Press, 323-390.

Pizlo, Z., Li, Y., Sawada, T., & Steinman, R. M. (2014). *Making a Machine that Sees Like Us*. New York: Oxford University Press.

Quilty-Dunn, J. & Mandelbaum, E. (2017). Against dispositionalism: Belief in cognitive science. *Philosophical Studies*, 175(9), 2353-2372.

Rahnev, D., & Denison, R. (2016). Suboptimality in perception.

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

Rescorla, M. (2014). Computational modeling of the mind: What role for mental representation? *Wiley Interdisciplinary Reviews: Cognitive Science*, 6(1), 65-73.

Reynolds, J. H., & Chelazzi, L. (2004). Attentional modulation of visual processing. *Annual Review of Neuroscience*, 27, 611-647.

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

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

Scholl, B. (2005). Innateness and (Bayesian) visual perception: Reconciling nativism and development. In *The Innate Mind: Structure and Contents*, eds. P. Carruthers, S. Laurence, & S. Stich. Oxford: Oxford University Press.

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.

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

Sober, E. (1976). Mental representation. *Synthese*, 33(1), 101-148.

Sperry, R. (1952). Neurology and the mind-brain problem. *American Scientist*, 291-312.

Sterelny, K. (1983). Mental representation: What language is brainese? *Philosophical Studies*, 43(3), 365-382.

Stich, S. (1983). *From Folk Psychology to Cognitive Science: The Case Against Belief*. Cambridge, MA: MIT Press.

Stich, S. (1992). What is a theory of mental representation? *Mind*, 101(402), 243-261.

Störmer, V., McDonald, J., & Hillyard, S. (2009). Cross-modal cueing of attention alters appearance and early cortical processing of visual stimuli. *PNAS*, 106(52), 22456-22461.

Taylor, J. (1962). *The Behavioral Basis of Perception*. New Haven, CT: Yale University Press.

Treue, S. (2004). Perceptual enhancement of contrast by attention. *Trends in Cognitive Sciences*, 8(10), 435-437.

Trivers, R. L. (2011). *The Folly of Fools*. New York, NY: Basic Books.

Tse, P. (2005). Voluntary attention modulates the brightness of overlapping transparent surfaces." *Vision Research*, 45(9), 1095-1098.

Turatto, M., Vescovi, M., & Valsecchi, M. (2007). Attention makes moving objects be perceived to move faster. *Vision Research*, 47(2), 166-178.

Wilbourn, M. & Kee, D. (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.

Wilson J., 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.

Zeman A., Obst O., Brooks K. R., & Rich A. N. (2013). The Müller-Lyer Illusion in a computational model of biological object recognition. *PLoS ONE*, 8(2), e56126. doi:10.1371/journal.pone.0056126.