Troubles with stereotypes for Spinozan minds

Bryce Huebner, Tufts University, Center for Cognitive Studies

Some people succeed in adopting feminist ideals in spite of the prevalence of asymmetric power relations. However, those who adopt such ideals face a number of psychological difficulties in inhibiting stereotype-based judgments. I argue that a Spinozan theory of belief fixation offers a more complete understanding of the mechanisms that underwrite our intuitive stereotype-based judgments. I also argue that a Spinozan theory of belief fixation offers resources for avoiding stereotype-based judgments where they are antecedently recognized to be pernicious and insidious.

1. Introduction

Being a feminist is hard, especially for a person who is on the privileged side of many asymmetric power relations. This is at least in part because numerous psychological barriers stand in the way of inhibiting misguided, insidious, and prejudicial stereotypical judgments.[1] But why is this the case? One answer to this question is painfully familiar: our perceptions are often organized on the basis of ideological structures that systematically disadvantage various groups of people. Such perceptions generate asymmetric power relations that are justified by way of post-hoc rationalizations, solidified as oppressive ideologies, and disseminated as true religion, unbiased education, and even as sheer entertainment. A sophisticated version of this story will explain why many people refuse to adopt feminist or egalitarian ideals. However, some people do adopt feminist and egalitarian ideals. In this paper, I focus on the difficulties that are faced by a person who recognizes that many of her stereotypical judgments are misguided and insidious, yet still finds many of her stereotypical judgments difficult to resist *as immediate intuitive judgments*.[2]

Authors Note: I would like to thank Jerry Doppelt, Luc Faucher, Randal Harp, Eric Mandelbaum (who introduced me to the idea that having a Spinozan mind has significant political implications), Susanne Sreedhar, Alison Wylie and two anonymous reviewers for Philosophy of the Social Sciences for their helpful suggestions on how to develop the central ideas of this paper.

[1] Similar difficulties are likely to be faced by people on the disadvantaged side of these power relations as well. There is a wealth of evidence regarding 'stereotype threat' (cf., Steele and Aronson 1995) that I believe is best explained in a way that parallels my discussion below. Unfortunately, addressing this topic would lead me astray of the main project of this paper; I thus leave that project for another paper.

[2] I am inclined to think that there are also similar stories to be told about the difficulties faced by those of us who wish to resist racist, heterosexists and classist stereotypes as well. However, the details of each of these stories, as well as the details of how various prejudices intersect will need to be worked out separately. In this paper, I merely propose one model of the psychological processes that underlie one sort of stereotypical judgment.

Focusing on the myriad barriers to the academic and professional success of women and visible minorities, many feminists have made a compelling argument for the claim that the relevant disadvantages are underwritten by processes that are systematic in nature. Such disadvantages call for redress at the level of social systems rather than at the level of individual psychologies. While I do not disagree, I argue that there is an important class of disadvantages that are grounded on implicit and unconscious biases that can only be understood by focusing on the structure of individual psychologies. This being the case, I hold that an adequate understanding of the possibility of redress for such disadvantages requires that we develop a more thorough understanding of the sorts of minds that we have, such that we are susceptible to numerous implicit and unconscious biases. I, thus, begin by bringing together the philosophical and psychological tools that are necessary to develop a plausible explanation of the formation of stereotype-based beliefs. With this account in hand, I turn to the cognitive barriers that are faced by those who intend to inhibit their stereotype-based judgments. Finally, I suggest a series of strategies that can be used to prevent the ironic rebound of stereotype-based judgments that occurs when a person finds herself judging in accordance with a stereotype that she intended to subvert all along.

2. TWO KINDS OF MINDS

In a recent study of the disadvantages generated by implicit stereotype-based judgments, Rhea Steinpreis and her colleagues (1999) hypothesized that whether a person was seen as 'hirable' or 'tenurable' could be modulated by changing the perceived gender of the name on a Curriculum Vitae (CV).[3] They sent a scientist's CV at either of two stages of her career to the members of various psychology departments in the United States. The first was the CV of someone who had secured her first tenure-track job directly out of graduate school; the second was the CV of the same person who had secured early tenure. On half of the CVs, the scientist's name was replaced with a paradigmatically male name (Brian Miller), on the other half her name was replaced with a paradigmatically female name (Karen Miller). Participants were then asked to evaluate the CV to determine whether the applicant would be hired (or tenured); what the applicant's starting salary should be; and, whether the applicant's teaching, research and service experience was adequate (Steinpreis et al 1999, 514). Steinpreis and her colleagues (1999, 522) found that although participants were no more likely to recommend

[3] Although I here address this recent study, this sort of inquiry into preferential hiring decisions on the basis of a CV has a long tradition. To my knowledge, the first study of this sort was carried out by Arie Lewin and Linda Duchan (1971). Although Lewin and Duchan (1971) found that responses trended toward a preference for male over female candidates, these results were not significant. I here focus on the results reported by Steinpreis et al (1999) because their analyses are more suggestive than other studies within this paradigm. Thank you to an anonymous reviewer for pointing me toward the earlier work in this paradigm.

tenuring a male than a female candidate, both male and female participants were more likely to recommend hiring a male over a female candidate. Moreover, *despite identical records* the research, teaching, and service contributions of male candidates were evaluated more positively than the same contributions of females. Finally, many participants claimed that they would need more evidence to demonstrate that female candidates had done their own work—though no similar claims were made about male candidates.[4]

There is no doubt that stereotype-based judgments play a significant role in producing these patterns of behavior and judgment. However, merely noting the presences of these effects offers little insight into the mechanisms that produce such behaviors and judgments. More importantly, in order to develop better strategies for avoiding the pernicious effects of stereotype-based judgments, we must first understand how these judgments are produced. Broadly speaking, there are two types of mechanisms that could facilitate the fixation of stereotype-based beliefs. Either stereotype-based beliefs are fixed through an effortful process of assessment and acceptance, or they are fixed non-consciously without any reflection whatsoever. Following the psychologist Daniel Gilbert (1990, 1991, 1993a, 1993b), I refer to the first as a 'Cartesian' theory of belief fixation and the second as a 'Spinozan' theory of belief fixation. Each of these theories of belief fixation offers an account of the mechanisms that produce stereotype-based judgments, an account of the difficulties faced by those who attempt to inhibit our stereotypical judgments, and an account of the techniques required for successfully overcoming the pernicious effects of unwarranted stereotype-based judgments. I argue that an adequate explanation of the psychological dimensions of implicit bias and stereotype-based judgment requires first understanding the mechanisms by which such judgments are fixed.

I begin with a brief sketch of the Cartesian theory of belief fixation (CBF). This theory has played a prominent role in philosophy, and it also seems to pervade commonsense.[5] Perhaps the clearest articulation of CBF is articulated in *Meditations*, when Descartes sketches a method for placing scientific knowledge on a firm and unshakable foundation. Roughly, Descartes argues that the scientist must withhold assent in any case where the truth of a belief is not absolutely certain.[6] This is, of course, the familiar skeptical strategy. However, if it is possible to adopt such a strategy, then the fixation of belief must be under the endogenous control of the person forming a particular belief.[7] This, however, requires the capacity to contemplate the truth of various ideas *prior to the formation of any belief whatsoever*. In brief, Descartes articulates a will-based theory of belief fixation that takes the following form. In the Cartesian mind, an idea (or in contemporary parlance, a representation) arises either through sensation or through the manipulation of other representations in the imagination. This is not to claim that there is no interaction between imagination and sensation; rather, the claim here is that there are two distinct processes that can, under some conditions, produce ideas independently of one another (though it is likely to be more commonplace for ideas to result from the integration of sensation-based processes and imagination-based processes). The idea that is produced in sensation or imagination is then passed forward to the will where it is evaluated for its fit with the world. This effortful process of evaluation and contemplation yields either an affirmation of the truth of the idea or a rejection of the idea as false. When the will affirms the truth of an idea, the output in the understanding is the belief that the idea is true (see figure 1).

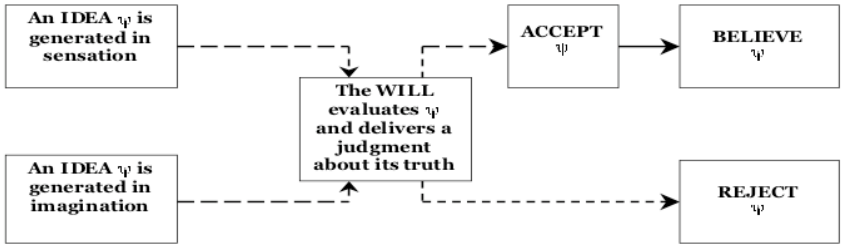

Figure 1: The Cartesian mind

According to CBF, whether a person has a particular belief is a function of the judgments that she makes about the truth of her representations. This does not mean that it is impossible to fall into bad habits for the evaluation of the truth of various representations. In fact, Descartes explicitly recognizes

[4] In a similar study conducted by Marianne Bertrand and Sendhil Mullainathan (2003), resumes were fabricated for non-academic jobs in Boston and Chicago. Half of the resumes were assigned a traditionally white name (e.g., Emily or Greg) and half were assigned a traditionally black name (e.g., Lakisha or Jamal). Bertrand and Mullinathan found that 'white applications' were approximately 50% more likely to yield requests for interviews than were 'black applications'. Moreover, this effect *was not* mitigated by considerations about the neighborhood in which the applicant lived.

[5] Were this not the case, convincing unsuspecting freshmen that Descartes skeptical worries are plausible would be far more difficult. There is also good reason to think that something like this view has played a prominent role in empirical psychology as well. Consider the following quote from Zimbardo and Lieppe (1991, 135; quoted in Gilbert et. al. 1993): "Learning requires that the audience pay *attention* to the message and, in turn, gain some *comprehension* of it, understanding the new beliefs it proposes. Then, if the message has compelling arguments, *acceptance* of its conclusion and a change in attitude will follow."

[6] Descartes' own view is far more sophisticated, and the Cartesian view of the mind that I articulate here is, thus, greatly simplified. However, the further details of Descartes proposal are not important for my purposes. My concern is merely to articulate a simple version of a will-based theory of judgment and belief fixation.

[7] For example, until the end of his *Meditations*, Descartes reports that he has withheld judgment about the truth of the idea that there is an external world, only allowing his will to affirm this idea after doing a lot of thinking about the nature of God and the nature of deception.

that withholding judgment in cases where we have formed similar judgments in the past requires a great deal of mental effort. However, if CBF is the correct model of our mind, then any person who exerts the necessary effort will be able to withhold judgment until she is certain of the truth of her representations. At this point, it is important to recall that CBF was developed in the service of resisting what Descartes deemed to be deeply entrenched ideas about the structure of the physical world. In essence, Descartes was searching for a methodology that would allow him to reject the dominant Scholastic ideology. Because CBF was hypothesized as a method for resisting ideology, it has important, though often unnoticed, political ramifications. Roughly, if the will plays a central role in the fixation of belief, then a person should have the capacity to withhold assent to any dominant ideology by a 'pure act of will'. Of course, it will often be difficult to withhold assent from insidious stereotype-based judgments in cases where we have made such judgments in the past. However, doing so will always be possible for a Cartesian mind. More importantly, this is true even when there are numerous social mechanisms that have the function of propagating and sustaining stereotype-based understandings of the world.

Despite its prominence, CBF is not the only game in town. In *Ethics*, Spinoza provides a simple and elegant alternative to this Cartesian view of the mind. According to Spinoza, people "believe themselves free because they are conscious of their own actions, and ignorant of the causes by which they are determined" however, "the decisions of the mind are nothing but the appetites themselves, which therefore vary as the disposition of the body varies" (Spinoza *Ethics* IIIp2s). As Spinoza saw things, rejecting Descartes' dualism about the mind also compelled him to reject CBF. After all, CBF removes the will from the causal order—making it a 'free cause of action' that resides outside of the natural world. In stark contrast to CBF, Spinoza argues that "there is no absolute, or free, will, but the mind is determined to will this or that by a cause which is also determined by another, and this again by another, and so to infinity" (*Ethics* IIp48). Spinoza took beliefs to be nothing more than the cognitive states produced in response to the stimuli that constantly impinge on a mind. In Spinozan terms, "The will and the intellect are one and the same" (*Ethics* IIp49c).

According to the Spinozan theory of belief fixation (SBF), ideas that are generated in sensation or imagination immediately yield a belief that the idea represents some feature of the world. Again, this is not to claim that there is no interaction between imagination and sensation, but only to note that there are two distinct processes that can, under some conditions, produce ideas independently of one another.[8] However, SBF, in stark contrast to CBF eliminates the will, and with it the capacity to withhold assent from the representations that arise through sensation and imagination (cf., *Ethics* IIp49S2). To put the point bluntly, *the Spinozan mind believes everything it reads*. However, if this is correct, then any sort of reflection that occurs concerning the truth of a belief can only take place after that belief has already been formed (see figure 2). To put this point another way, reflecting upon and rejecting a belief can only take place when it generates some sort of noticeable conflict in a person's overall understanding of the world.

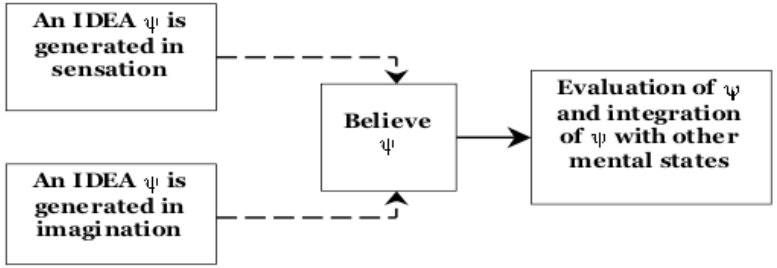

Figure 2: The Spinozan mind

With this view of belief fixation in mind, Spinoza also recommends a strategy for resisting ideology: avoid exposure to ideology; and if you are exposed to ideology, exorcise your ideological beliefs as soon as possible. According to SBF, resistance to insidious stereotypes is likely to be a far more difficult task. After all, the Spinozan mind makes stereotype-based judgments *immediately* as it is presented with the words and images that flash across a television screen, or that are printed in *Cosmo* or *Maxim*; it forms beliefs on the basis of every claim advanced in a philosophy department and every claim contained in an academic paper. This means that the social mechanisms that propagate and sustain insidious stereotypes will make it exceedingly difficult to notice all of the stereotype-based beliefs that we acquire. For, it is only when there is a noticeable conflict in a person's overall understanding of the world that beliefs will be evaluated and excised. But for the person who is on the privileged side of many asymmetric power relations, such noticeable conflicts are likely to be far too rare. Social-ideological mechanisms are likely to *compel* Spinozan minds to acquire numerous stereotypical beliefs; and, if belief fixation is Spinozan in nature, *it's going to be hard to exorcise all of these insidious stereotypical beliefs* because denying them will require an *active* recognition that already formed beliefs are false. More importantly, this process of exorcising insidious stereotypical beliefs is likely to be made much more difficult because we are continually

[8] This model of SBF, just as the model of CBF developed above, has been simplified for ease of presentation. There are, for example, likely to be important feedback loops between sensation and imagination. The claim here should be quite familiar. Ideas that are produced by sensation can often be enlivened by the imagination, and ideas that are produced in the imagination can often modulate our perceptions. Both Descartes and Spinoza recognized this fact. However, this does not rule out the possibility that there are inputs to the system dedicated to belief fixation that arise exclusively in the imagination or in sensation.

bombarded with ideology that will further solidify our stereotype-based judgments.

3. SPINOZAN MINDS AND AUTOMATIC PROCESSES:

Despite its initial intuitive plausibility, CBF is ill-supported on both empirical and conceptual grounds. Of course, a defense of the Spinozan alternative as a complete theory of the mind is beyond the scope of this paper. However, a striking set of results that have been collected by social psychologists over the past twenty years suggests that our behavior is often produced automatically in response to apparently innocuous stimuli, in the absence of any act of will *whatsoever*. This is not to deny that conscious processes and deliberate choices sometimes play a role in the human mind; rather, the claim that is supported by the relevant psychological data is that there are numerous non-conscious processes of belief fixation that play an integral role in structuring our judgments.

The most intriguing evidence for this claim has been collected by John Bargh and his colleagues. For example, people who are implicitly primed with terms stereotypically associated with the elderly (e.g., worried, old, lonely, Florida) walk more slowly to an elevator (Bargh et. al. 1996). Moreover, people who are presented with such implicit primes show a decrease in performance on standard memory tasks (Bargh et. al. 1996). Put briefly, thinking about the elderly makes you more forgetful. More troublingly, people who are subliminally primed with the face of an African-American male become more hostile when the computer that they are working on fails during a tedious task (Bargh et. al. 1996). Similar results proliferate throughout social psychology. People who are primed with 'professor' outperform those who are primed with 'soccer hooligan' at games of trivia (Dijksterhuis and Van Knippenberg 1998). In an ultimatum game, people who are in a room with a backpack are more likely to cooperate, while those who are in a room with a briefcase are more likely to be competitive (Kay et al 2004). Finally, people who are shown a picture of a library and instructed to go to the library after the experiment speak more softly *during* the experiment (Aarts and Dijksterhuis 2003).

Though these results cannot *prove* that non-conscious processes direct our behavior, they are suggestive. However, it remains an open question just how robust these non-conscious and automatic processes of belief formation actually are. In order to spell out a robust version of SBF, judgments of truth and falsity, as well as immediate behavioral responses, must all be affected by non-conscious and automatic processes of belief fixation. After all, even a Cartesian mind will be susceptible to differences in the organization of the information with which it is presented. However, while a Cartesian mind has the capacity to recognize immediate and intuitive responses as false before they are encoded as beliefs, a Spinozan mind does not. But, unfortunately for the Cartesian, a wealth of psychological evidence suggests that we cannot just reject all of our beliefs.

To begin with, numerous psychological studies have demonstrated that repeated exposure to a claim, even where it is explicitly labeled as false, will increase the likelihood that it will be believed to be true (cf., Gilbert et. al. 1993 for a review of this data). Of course, the Cartesian could object that this was the result of sheer laziness. After all, we clearly tend to use a heuristic that says 'believe the things that you are told repeatedly'—and, this doesn't seem like a bad strategy overall. However, another set of findings suggests conditions where the capacity to unbelieve something that you are told can be unconsciously overridden (Gilbert et al, 1993). One condition where such an override takes place is during an increase in cognitive load (i.e., an increase in the amount of cognitive work that is required for executing a particular task).

To demonstrate the truth of the claim that you can't unbelieve everything that you read, Gilbert and his colleagues (1993) asked people to read aloud as the evidence from crime reports scrolled across a computer screen. These people were told to take care to distinguish the facts of the case, which were always printed in black, from irrelevant information that belonged to a different report, which were always printed in red. Unbeknownst to the participants, the irrelevant information exacerbated the severity of the crime. In order to generate an increased cognitive load, half of the participants engaged in a simultaneous number-search task while they read irrelevant information. After reading the entire crime report, participants were asked to consider the facts of the case and to recommend a prison term for the criminal. The results were quite shocking: people under an increased cognitive load reliably recommended prison sentences that were nearly twice as long in the presence of irrelevant information exacerbating the severity of the crime. In order to strengthen the case for SBF, Gilbert and his colleagues also asked their participants to engage in a recognitional memory task after having settled on a prison term. While they found no difference in performance on this task for the true statements, they did find that people who had been subjected to an increased cognitive load were more likely to remember false statements *as true*.

As with many results reported within the cognitive sciences, these data suggest that the human mind is deeply fragmented.[9] It operates in terms of two distinct sorts of processes (cf., Chaiken and Trope 1999). First, it includes non-conscious and automatic processes (hereafter, *Type-1 processes*) that yield behavior and belief without conscious processing or decision-making.[10]

[9] For a philosophical defense of this claim, see Andy Egan (forthcoming) and the papers cited therein.

[10] In advancing this data as evidence for the truth of SBF, I am adopting the claim that the modulation of behavior that occurs in these tasks provides evidence for a corresponding modulation of belief. Of course, these beliefs are non-conscious, and most of the models that have been developed in order to account for these phenomena hold that these beliefs cannot be made conscious. I see no compelling reason to think that consciousness is a necessary condition on being a belief. However, the argument in this section holds even on the weaker assumption that numerous behaviors are produced by non-conscious operations.

Second, the mind also includes deliberative processes (hereafter, *Type-2 processes*) that operate within the realm of consciousness and rationality.[11] The data canvassed thus far, as well as a wealth of data from the social sciences, suggest that many of our stereotype-based judgments result from the operation of *Type-1 processes*.[12] Superficial features associated with race, gender, age, and numerous other categories have the capacity to produce stereotype-based judgments in the absence of any sort of conscious reflection. Given that this is the case, one might wonder *why* we would be likely to categorize on the basis of stereotypical features.

In some cases, the explanation of such categorization is clear. After all, if we had to work to convert the deliverances of every perceptual episode into a belief, "we'd be slower and less reliable in our uptake of perceptually given information, especially when some of our cognitive resources" were otherwise occupied (Egan forthcoming, 19). To put the point another way, if a *Type-2 process* had to be engaged in order to allow a person to avoid an angry tiger, human beings would not have been successful organisms, evolutionarily speaking. Such stories are nice, where they apply. However, while there are cases where we are better off in our rapidly changing and dangerous environment because perception immediately generates belief, the story is more complicated when it comes to the information required for automatically encoding many of the stereotypical features that we encode. For his part, Spinoza (*Ethics* IIIp46) offers a simple proto-associationist suggestion for the automatic categorization of a person on the basis of stereotypical features. He claims that if a person is:

> affected pleasurably or painfully by anyone, of a class or nation different from his own, and if the pleasure or pain has been accompanied by the idea of the said stranger as cause, under the general category of the class or nation: the man will feel love or hatred, not only to the individual stranger, but also to the whole class or nation whereto he belongs.

Spinoza's suggestion is obscure. However, the key point of import for my purposes is that an adequate understanding of the psychology of a stereotype-based judgments requires this intuitive and unreflective capacity to categorize.

Humans are social animals; and, as such, it would be strange if we lacked the psychological capacity to categorize on the basis of group membership.[13] While we do not currently have an adequate account of how group membership is determined psychologically, we should immediately recognize that "the ability to understand new and unique individuals in terms of old and general beliefs is certainly among the handiest tools in the social perceiver's kit" (Gilbert and Hixon 1991, 509). The process of stereotype-based categorization allows us to make simplifying assumptions about our social world. Moreover, it allows us to make immediate and unreflective judgments about the categories to which a person belongs so that we can rely on numerous representations that are already present in long term memory (cf., McRae and Boddenhausen 2000, 95). Finally, relying on stereotypical representations provides a set of background assumptions on which perceptions of within-category difference can be rapidly processed. That is, by making simplifying assumptions about the categories to which a person belongs, we do not have to focus our attention on the familiar, but can instead focus on encoding unexpected features that will be more likely to be informative for navigating our social world (cf., McRae and Boddenhausen 2000, 106).

Of course, we should still wonder why we categorize on the basis of gender. In fact, although it is widely recognized that we rapidly and automatically classify on the basis of gender, there is little consensus within the cognitive sciences concerning the mechanisms that underwrite immediate and automatic categorization.[14] Some cognitive scientists appeal to the fact that members of a sexually dimorphic species must rapidly categorize in terms of prospective mates; others appeal to our rapid classifications on the basis of body movement and gait (cf., Cutting and

[11] Building on the evidence for a dual-process model of human psychology, Peter Caruthers (2007) has argued for a version of the full-blown Spinozan hypothesis (though he does not explicitly recognize it as such) that the will is an illusion. I find his argument quite compelling, though I will not rehearse it here since I only need the weaker claim that some stereotype-based judgments arise in the absence of a conscious or deliberate choice.

[12] See Wegner and Bargh 1998 for a review of the data on the automatic activation of stereotypical representations. There is also an affinity between my discussion of *Type-1 processes* and the discussion of schemas by Taylor and Fiske (1991) and Valian (1999).

[13] Mark Schaller and Steven Neuberg (in press) review a wealth of evidence suggesting that the capacity to classify people in terms of 'us' versus 'them' is a primitive aspect of human cognition. They also demonstrate that people have the capacity to detect a variety of threats that might come from members of an out-group. Of course, it would be startling to find that this was a capacity that we lacked. After all, the capacity to treat members of out-groups seems to be present quite a ways down the phylogenetic tree (at least down to ants and bees!). As many philosophers and cognitive scientists have argued, our affective responses to various sorts of stimuli can be used to coordinate adaptive responses to various sorts of environmental threats (cf., Griffiths and Scarantino 2005). On the basis of this assumption, Luc Faucher, Eduoard Machery, and Daniel Kelly (in prep) review a wealth of data suggesting that we represent different social groups as posing different sorts of threats to 'us'; these different threats to evoke different sorts of affective responses that generate differential prejudicial attitudes to various groups of people. For example, people who are afraid of catching a disease are more likely they are to exhibit xenophobic attitudes toward unfamiliar groups. Moreover, when a white person believes that the world is a dangerous place, and she is placed in a dark room, her stereotypical judgments about blacks increase. None of this, of course, requires anyone to adopt an evolutionary picture of race *as such*. At minimum, it might be that we possess an innate in-group/out-group schema that is elaborated culturally in racialized terms. I intend to avoid all questions concerning the nativism of racial representations in this paper.

[14] In the latter case, however, it must be noted that the evidence reported by Bülthoff and Newell (2004) shows that this sort of sex-based categorical perception occurs only for familiar faces, and not for unfamiliar faces.

Caferelli 1977; Johnson and Tassinary 2005) or facial features (cf., Bülthoff and Newell 2004). However, regardless of the perceptual mechanisms that facilitate the categorical perception of gender-based stereotypes, it must also be acknowledged that there are multiple cultural influences that facilitate the immediate perception of gender-based categories. Even if we are innately disposed to categorize on the basis of *sex*, there is little doubt that we *are not* innately disposed to make judgments about intelligence, mathematical competence, of intellectual ability on the basis of *gender*.[15] Any theory of gender-based categorization must, then, recognize the effect of ideological structures that generate and propagate insidious stereotypes; this ideology leads us to associate various social roles with irrelevant, supposedly sexually dimorphic cues. This said, while we know little about the psychological mechanisms that facilitate categorization on the basis of gender-based stereotypes, it is safe to assume that such categorization occurs because of the operation of a *Type-1 process*. That is, information about a person's gender is typically processed before any conscious deliberation occurs.

Of course, this should not come as a surprise; however, this fact has important consequences for minds like our own. Although there may be *some situations* where rapidly categorizing on the basis of supposedly sexually dimorphic cues would be a good strategy for navigating a social world, possessing such a rapid and unreflective belief forming process is deeply problematic in our world. We rely on this process of categorization in cases where its outputs are utterly irrelevant to the decisions that must be made. Of course, some people consciously affirm the stereotype-based judgments produced by these *Type-1 processes*. However, for those who acknowledge that many stereotype-based judgments are both misguided and unjustifiable, the important question to ask is whether egalitarian *Type-2 processes* can be recruited to override a stereotype-yielding *Type-1 processes*.[16]

4. SPINOZAN MINDS AND STEREOTYPE REBOUND:

Spinoza drew a distinction that approximately maps the distinction between *Type-1* and *Type-2 processes*; he also saw the automaticity of thought as an *ethical* problem. He argued that insofar as beliefs are generated automatically in response to whatever stimuli happen to arise, we are *necessarily passive* (*Ethics* IIIP56D). That is, as long as beliefs are generated automatically, we have no control over the sorts of lives that we lead. It would be nice, then, if there were an easy strategy for preventing problematic *Type-1 processes* from generating behavior. However, overriding *Type-1 processes* is more difficult than one might hope.

Part of the explanation for this difficulty lies in our failure to understand the mechanisms that drive the processes that we are attempting to override. Because we are, by definition, intimately aware of the operation of *Type-2 processes*, but completely unaware of the operation of many *Type-1 processes* that generate belief and behavior, we assume that deliberate thought exhausts our mental lives. This misguided focus on conscious thoughts allows us to *say* that we do not have sexist beliefs without inhibiting any *Type-1 processes* that may continue to generate stereotype-based judgments or behaviors. We are inclined, just as Descartes was, to believe that we have the capacity to withhold our assent from any beliefs that happen to crop up. However, there are some conditions under which suppressing stereotype-based judgments poses a nearly insurmountable task for Spinozan minds like our own.

The difficulties inherent in suppressing thoughts are nicely encapsulated in a task proposed by Fyodor Dostoyevsky (1863/1997, 49): Don't think of a polar bear. If you try to do this, "you will see that the cursed thing will come to mind every minute". Recalling this example, Daniel Wegner and his colleagues (1987) trained participants to report on their stream of consciousness in an experiment on thought-suppression. They found that those people who suppressed thoughts about white bears *had more thoughts about white bears in a later phase of the experiment* than did people who

[15] Of course, there are myriad ways of drawing a distinction between sex and gender. I worry that none of them will be capable of drawing an exclusive distinction between these two classes. However, my intention is merely to call attention to the fact that although there are some biologically-based categories that an evolutionary story may be well suited to explain, similar explanations are likely to fail for more socially structured processes of categorization. At this point, I am inclined to think that part of the story about why we have the gender-based stereotypes comes in the form of *perceived* differences in informational specialization (cf., Wegner et. al. 1991). These differences are, of course, culturally entrenched. However, there is some reason to think that gender-based categorization allows us to make simplifying assumptions about the sorts of information that a person is likely to know, assigning plausible domains of expertise on the basis of her coarse-grained morphological properties. At present, there is some support for this hypothesis in the literature on transactive memory; however, it is far from clear how robust a role this data will play in explaining why we classify on the basis of gender.

[16] An anonymous referee has raised the following worry. Even if we can draw a distinction between processes *that seem* intuitive and unreflective and processes *that seem* deliberative, this cannot establish the existence of genuinely deliberative processes. That is, our minds could be exhaustively constituted by *Type-1 processes*, some of which masquerade as *Type-2 processes*. I concede that this might be the case. However, I hold that even if all of the processes at play in our minds are *Type-1 processes*, this does not rule out the possibility of deploying more reflective processes in the evaluation of various beliefs for their fit with other beliefs that were held antecedently. The most promising distinction to be drawn in articulating this version of SBF is by recognizing a distinction between person-level processes and sub-personal processes. Recent cognitive science has demonstrated that our minds consist of a variety of sub-personal mechanisms that are dedicated to the fixation of beliefs for relatively restricted domains of inputs. However, humans (and perhaps we are unique in this respect) have the capacity to generate person-level conceptual representations that are decoupled from their input systems and that can be redeployed across numerous domains in categorizing other sorts of experiences. According to the weakest reading of SBF, these person-level representations can be mobilized in order to search for inconsistencies in a person's overall cognitive economy, thereby modulating the conscious, person-level representation of the structure of the world. This is not, of course, to say that such mechanisms are free of constraints. Instead, the claim only need to be that the Spinozan mind contains a system that can 'reflect' on the structure of its overall set of beliefs, excise those beliefs that conflict with more centrally held beliefs, and that this process can be under endogenous control at the person-level rather than at the non-conscious level. This much, I think, should be perfectly acceptable even to the hard-determinist.

had did not suppress their white bear thoughts. Wegner and his colleagues have found similar results in a number of domains. On the basis of these results, Wegner has argued that thought-suppression often has the unintended ironic effect of causing people to obsess about the thought that they were trying to suppress. Thought-suppression can thus cause a thought to rebound.

SBF predicts this ironic rebound effect in conditions where only *Type-2 processes* are suppressed but both *Type-1* and *Type-2 processes* are operative in producing judgments. After all, suppressing one's conscious thoughts about white bears does not necessarily suppress all of the processes that might continue to generate white bear thoughts. In fact, the suppression of *Type-2 processes* is likely to engage a *Type-1* monitoring process dedicated to detecting the thought that is to be suppressed (cf., Wegner et. al. 1987). As a result of the operation of this *Type-1* process, the suppressed thought will continue to be active in the agent's overall cognitive economy. Thus, in order to suppress white bear thoughts, a system that utilizes a white bear representation must continue to operate, and this mechanism presents serious difficulties for the process of 'weeding out' all of the thoughts that were intend to be suppressed. The problem here is that prior to any conscious deliberation, more instances of the thought that is to be suppressed are going to be incorporated into the agent's overall cognitive economy.

On the basis of SBF, we should, thus, expect stereotype-based judgments to exhibit a similar ironic rebound effect; and, this is precisely what we find. Building on Wegner's results, Neil Macrae and his colleagues (1994) asked people to view a photograph of a skinhead and then to write a story about a day in his life. Half of the people were told to suppress stereotypical thoughts about skinheads. In a later phase of the experiment, one-third of the participants were asked to write a second story about a skinhead from another photo; those participants who had initially suppressed skinhead stereotypes in the first phase of the experiment used more stereotypes in writing their second story. Another third of the participants were taken to a room where they were told that they would meet the skinhead about whom they had just written their story. They were told that the skinhead had stepped out to use the restroom, and they were asked to take a seat in a room that contained a number of chairs, and a backpack and jacket that purportedly belonged to the skinhead. In this condition, participants who had suppressed stereotypical judgments in the first phase of the experiment chose a seat that was farther away from the skinhead's belongings than did those who had not suppressed their stereotypical thoughts. Finally, some participants performed a lexical decision task; in this case, those who had suppressed stereotypical judgments in the initial phase of the experiment responded more quickly in classifying stereotypical words as words.[17]

[17] Similar results have been found for stereotypical judgments about 'male construction workers', 'yuppies', and 'politicians'.

These results suggest that some stereotype-based judgments tend to be strengthened when a person suppresses her conscious stereotype-based thoughts. However, there is some reason to think that although *some* stereotypes rebound, the most pernicious ones will not. After all, there are ways of avoiding the ironic results of thought-suppression. In their initial experiments on thought-suppression, Wegner and his colleagues (1987) found that the ironic results of suppressing thoughts about white bears could be prevented by focusing on thoughts about red Volkswagens. As Spinoza suggests (*Ethics* IVP7), it is possible to restrain the output of a *Type-1 process* by focusing on "an affect opposite to, and stronger than, the affect to be restrained". Where pernicious gender-based stereotypes are concerned, we might hope that the social pressure to think in egalitarian terms would yield strategies of focusing on thoughts that are incongruent with stereotype-based judgments. Thus, although people do not have effective coping strategies for suppressing stereotype-based judgments about skinheads, they might have effective strategies for suppressing more pernicious stereotypical judgments that are grounded on more socially significant categories.

On the basis of this sort of assumption, Margo Monteith and her colleagues (1998) have provided evidence that stereotype-based judgments about homosexuals do not rebound for low-prejudice heterosexuals.[18] To show that this was the case, Monteith and her colleagues (1998) utilized the two-story condition from McRae et. al. (1994) and asked people to write a story about the day in the life of a homosexual couple. Monteith and her colleagues found that stereotype suppression resulted in ironic rebound only for high-prejudice participants, suggesting that low-prejudice individuals have the capacity to avoid stereotype rebound by using egalitarian beliefs to inhibit the encoding of *Type-1* stereotype-based judgments as beliefs. High-prejudice individuals, on the other hand seem to have little motivation to counter their immediate stereotypic reactions and so do not generate easy replacements for their problematic stereotypical thoughts. If only things were this easy!

Although egalitarian attitudes have the capacity to significantly reduce the ironic effects of stereotype suppression, things are more difficult when we turn to role of egalitarian attitudes in modulating stereotype-based judgments in the world of our everyday experience. There is, after all, reason to think that egalitarian attitudes are fairly prevalent in academia—or are at least that they are becoming more so. However, such egalitarian attitudes are not by themselves sufficient to obviate gender-based inequalities in the academy. As Virginia Valian (1999, 2) argues:

> Conscious beliefs and values do not, however, fully control the operation of nonconscious schemas. Egalitarian beliefs may help, but they do not guarantee accurate, objective, and impartial evaluation and treatment of

[18] Participants were rated for overt prejudice against homosexuals on the basis of the Heterosexual Attitudes Toward Homosexuals (HATH) scale (Larsen, Reed, and Hoffman 1980).

> others. Our interpretation of others' performance are influenced by the unacknowledged beliefs we all—male and female alike—have about gender differences.

Following Valian, there is a straightforward explanation for why egalitarian attitudes are insufficient to facilitate successful stereotype suppression. Although we can rid ourselves of some stereotypical judgments by replacing them with egalitarian beliefs, we are not yet aware of all of the *Type-1 processes* that yield stereotype-based judgments. While adopting egalitarian attitudes is helpful for getting rid of stereotype-based judgments that we are aware of, it can only help where we know that we are operating on the basis of such judgments.

Hypothesizing that some *Type-1 processes* would not be overridden by the adoption of egalitarian attitudes, Sei Jin Ko and her colleagues (in press) asked people to write a story about a day in the life of a male and a female shown in two different photos. In a subsequent, purportedly unrelated task, participants were then asked to read a brief stereotypically feminine story, and half were asked to suppress their stereotypical thoughts.[19] Participants were then asked to rate the probability that each of a series of voices (previously rated as paradigmatically male or female) belonged to the author of the story that they had just read. Although all of the participants succeeded in suppressing categorical judgments about gender, participants who had suppressed stereotypical judgments tended to rely more heavily on feature-based stereotypes about vocal femininity in determining who had written the story than did controls. That is, participants who had suppressed stereotypical judgments were far more likely to say that the author of the story must be the person who sounded the most feminine. This suggests that although some stereotype-based judgments can be avoided by adopting egalitarian thoughts, modes of categorization based on gender-typical features may continue to operate without us ever noticing this.

Unfortunately, things get even worse. If SBF is correct, then suppressing even the most obvious sorts of stereotype-based judgments requires an effortful and deliberate attempt to disbelieve stereotype-based judgments that have already been automatically generated. In a Spinozan mind, the only way in which a problematic *Type-1 process* that yields stereotype-based judgments can be countered is by engaging a *Type-2 process* that can excise the stereotype-based judgments that have already been generated. But, if this is the case, then an increase in cognitive load should make it exceedingly difficult to suppress a stereotype even where it is recognized as insidious. As the Spinozan theory of the mind predicts, stereotype rebound occurs even for socially sensitive stereotypes when cognitive load is increased. For example, people who are told not to be sexist in completing ambiguous sentence stems such as "women who go out with a lot of men are…" are more likely to complete these stems on the basis of sexist generalizations in those cases where they were faced with increase in time pressure (Wegner et. al. 1993, cited in Wegner 1994). Similarly, when asked to rapidly classify sexist and nonsexist statements as true or false, people who are instructed not to be sexist tend to judge sexist statements true and egalitarian statements false more quickly than they are able to judge sexist statements false and egalitarian statements true (Wegner 1993). Absent the instruction not to be sexist, there is no difference in the latencies of these judgments. Moreover, and this is the important effect, these results are not modulated by overt prejudice!

Such cases of stereotype-rebound as they occur in the lab are interesting. But, there are further important implications for this effect in the world of our everyday experience. While there are numerous situations in which stereotypes can rebound, I shall focus only on the rebound of stereotypes in academic settings. First, imagine the paradigmatically masculine male who is introduced to feminist ideals in the context of an introductory course on feminist philosophy. Suppose that he understands the problematic asymmetries generated by traditional gender-roles—at least when he is thinking about them inside the classroom. However, despite his best efforts to avoid making stereotype-based judgments, he is unsuccessful. When he leaves the classroom, his understanding rapidly dissipates, and he continues to believe that the asymmetric power relations instantiated *in his life* really do map gender-based differences. He thereby continues to facilitate the asymmetric power relations that he initially recognized as problematic.

Second, imagine the male philosopher who adopts feminist ideals, but who continues to find himself with numerous stereotype-based beliefs. He recognizes that there are no *intrinsic differences* in the capacity of males and females to do philosophy well. He even recognizes that although he often makes stereotype-based judgments about various groups of people, many of these judgments are radically misguided; so, he attempts to suppress stereotype-consistent beliefs whenever they occur. However, he continues to judge that the questions asked by his female colleagues in department colloquia demonstrate a lack of the analytical capacities he finds in his male colleagues. Although, he recognizes that there are no category-based differences between the abilities of men and women to do philosophy well, he judges that *his* female colleagues lack the qualities necessary for doing philosophy well.

Finally, imagine the philosopher committed to resisting stereotype-based judgments in his evaluation of job dossiers. Because he is aware of the tendency to make immediate judgments that will be grounded in stereotypical representations of people, he recognizes that he must be careful not to allow these stereotype-based judgments to structure his decisions concerning whom to interview. However, although he finds no significant differences in the quality of the written work provided by male and female

[19] For example, some participants read the following story: "As an elementary school teacher, I like to create an environment where students learn to cooperate and build self-confidence. An essential part of doing this is not to have favorites, but rather to give more care and attention to the children who are more shy and reticent. I make myself available even outside of the classroom if any one of them should need my help" (Ko et. al. in press, 3).

candidates, he still finds that his subsequent interactions with female candidates suggest that they are not as philosophically sophisticated as their male competitors and he finds himself unsure of whether they will be good colleagues or advisors.

Such failures in suppressing stereotype-based judgments are unlikely to come as a surprise to anyone who has actively engaged in resisting such judgments. In fact, many feminist activists have canvassed these failures in attempting to rectify the pernicious effects of gender and race-based stereotypes in the academy.[20] However, I suggest that adopting the dual-process theory of the mind that I have been drawing on in order to defend SBF helps to explain *why* people who wish to suppress insidious stereotype-based judgments often fail. I argue that a more complete understanding the relative contributions of *Type-1* and *Type-2 processes* in the production of stereotype-based judgments offers better insight into the mechanisms that produce the rebound of suppressed stereotypes. In the remainder of this paper, I turn briefly to these three real-world cases and suggest that this dual-process theory of the mind can help us to understand why it is so difficult for a person who seems to have a desire to inhibit her stereotypical judgments to successfully do so. I then conclude with some suggestions about how to recruit the psychological and social resources that are required in order to facilitate successful resistance to insidious stereotype-based judgments.

5. SPINOZAN MINDS AND THE AVOIDANCE OF STEREOTYPE REBOUND?

First, consider the masculine male student who, *in the classroom*, understands the problems with asymmetric gender-based power relations, but who continues to live his life as though asymmetric gender-based stereotypes are veridical. This person is capable of suppressing his stereotype-based judgments by engaging a *Type-2 process* that outputs egalitarian beliefs when he is in a social situation that he sees as obviously requiring him to do so. However, numerous *Type-1* and *Type-2* processes continue to generate stereotype-based judgments that he does not realize he should be suppressing. He is, for example, unlikely to entertain alternatives to traditional gender-roles (cf., Toller et al, 2004). Moreover, because he sees himself as masculine, he is likely to treat 'feminist' as indicative of *femininity*, and thereby, to be avoided. The important thing to notice about this case is that masculine males often believe that they have more to lose by adopting an egalitarian worldview. This precludes one of the more straightforward strategies for suppressing stereotype-based judgments: focusing on egalitarian thoughts. Because the mind of the masculine male is likely to have numerous *Type-1* and *Type-2* processes that yield stereotype-based judgments, the masculine male often experiences the most straightforward sort of ironic stereotype rebound.

Second, consider the male philosopher who adopts feminist ideals, but who continues to have numerous stereotype-based beliefs. Although he, unlike the paradigmatically masculine male, is likely to have developed some effective strategies for utilizing egalitarian beliefs in order to resist category-based stereotypes, this strategy is likely to fail in a number of cases because he is unaware of all of the ways in which *Type-1 processes* are producing stereotype-based judgments. Although he exerts the effort required to overcome many of the stereotype-based judgments that are produced by *Type-2 processes*, he fails to inhibit the stereotype-based judgments produced by *Type-1 processes*. Thus, suppressing the judgment that men are more intelligent than women, is likely to result in an obsessive focus on the 'lack of analytical capacities' that he finds in his female colleagues (NB: this is a lack that he would also find in many of his male colleagues *if he were looking*). This is, of course, precisely the sort of situation in which many academics are likely to find themselves (cf., Valian 1999).

Finally, recall the philosopher who is actively committed to resisting insidious stereotypes in evaluating job dossiers. He is aware of his tendency to make intuitive stereotype-based judgments. He also recognizes that he must be incredibly careful not to allow these stereotype-based judgments to structure his decisions about who to interview. However, having suppressed his initial stereotype-based judgments, he may find that although there are no significant differences in the quality of written work provided by the male and female candidates, subsequent interactions with female candidates still suggest less philosophical sophistication than interactions with male candidates. That is, although he adopts egalitarian views, and although he has the capacity to inhibit the formation of stereotype-based judgments produced by *Type-1 processes* by ensuring that they are made obviously incongruent with the rest of his egalitarian thoughts, when he is under an increased cognitive load he will still be likely to experience the ironic effects of stereotype suppression. The reason for this is that preventing stereotype-based judgments from being integrated into his cognitive framework requires him to engage in the effortful *Type-2 process* of excising beliefs that he recognizes as insidious. However, when he cannot dedicate the necessary attention to preventing these beliefs from being integrated into his overall cognitive economy, the *Type-1 processes* that are producing stereotypical judgments continue to operate in such a way that they generate behaviors that he would, in a reflective moment, be unwilling to endorse.

The important thing to keep in mind about this last case, however, is that increases in cognitive load are likely to occur in precisely the sorts of cases where we most want to inhibit insidious stereotype-based judgments. It is not only in cases where a person has to remember a lot of things about a potential job candidate where we find such an increase in cognitive load, but also cases in which a decision has to be made quickly, or under the pressure

[20] For example, the programs designed by the ADVANCE program at the University of Michigan, especially the those within the comprehensive workshop designed by the STRIDE committee to address unconscious bias, are precisely the sorts of strategies suggested by the mechanisms that I address in this paper.

of a lot of distractions that require an increase in attention to what a person is saying or doing. Given that we live in a world filled with distractions, it is likely that most of our lives will continue to be governed by the operation of *Type-1 processes* that generate beliefs that we do not want to endorse (cf., Mandelbaum in prep). This, however, brings the ethical implications of having a Spinozan mind front and center.

As Spinoza argued, there are numerous cases in which conflict arises between the beliefs generated by *Type-1 processes* and our more reflective judgments. Unfortunately, because we are typically unaware of the operation of these *Type-1 processes*, they often subvert the *Type-2 processes* that yield reflective desires to avoid problematic judgments. Because most of us find it unbearable to have obviously contradictory beliefs; so, we tend to justify stereotype-based judgments produced by *Type-1 processes* in order to appease cognitive dissonance (cf., Festinger et. al. 1956). This is not to say that we take our stereotype-based judgments to be justifiable *as such*. However, because we take ourselves to act on the basis of reasons, we often seek out reasons that will justify the judgments that we have already made. However, for a Spinozan mind, this tactic of appeasing cognitive dissonance is counterproductive where the *Type-1 processes* are problematic! The beliefs formed in order to justify our prejudicial judgments are likely to yield more beliefs that are consonant with the insidious stereotype that we are trying to subvert, thereby producing more problematic beliefs that themselves need to be exorcized.

At this point, however, it becomes clear why we are lucky to have a psychology that consists of both *Type-1* and *Type-2 processes*. As Andy Egan argues, if our psychology consisted of only one sort of process, we would be in much worse shape. For as soon as the deliverances of a *Type-1 system* were produced, they would "crowd out all of your hard-won confidence of the mechanism's unreliability, and all of your confidence in the evidence that convinced you of its unreliability in the first place" (Egan forthcoming, 22). However, because we possess both *Type-1* and *Type-2 processes*, it is *possible* for us to mobilize *Type-2 processes* in order to override the *Type-1 processes* that produce beliefs that we cannot reflectively endorse.

Reflection can occur, if it occurs at all, only at the level of *Type-2 processes*. There are, however, numerous questions that arise in determining which sorts of immediate and intuitive judgments ought to be revised and which ought to be retained. I suggest, however, that it is an open and empirical question when the outputs of *Type-1 processes* are adequate and when they ought to be revised by utilizing more reflective strategies to modulate these immediate and intuitive judgments. To put the point succinctly, any theory that allows for the operation of non-conscious processes of belief fixation will have to acknowledge that we are likely to be susceptible to cognitive illusions. To make the theoretical importance of this claim clear, consider an analogy from vision science. The relatively informationally encapsulated structures of the visual system yield representations of lines of different lengths in the Müller-Lyre illusion and representation of differences in shade in the Craik-Cornsweet-O'Brien illusion. This is true despite the fact that the two lines are the same length and the blocks are the same shade. On the Spinozan model of the mind, we can learn not to trust the immediate deliverance of vision even though vision will automatically produce a belief that the lines are different lengths and that the blocks are different colors. Similarly, where we find that *Type-1 processes* yield stereotype-based judgments that fail to track the actual structure of the world, we can deploy *Type-2 processes* to excise our stereotype-based beliefs.

This possibility, of course, turns on adopting the deliverances of scientific and ethical inquiry into the structure of the world as *more central* to our overall belief set than out immediate intuitive judgments. This means that even our deliberative and reflective judgments will always be open to further ideological contamination. This is an unfortunate situation for us to be in. However, there is no Archimedean standpoint for making such scientific and ethical judgments about the world. So, our best hope is to start from the most epistemologically promising standpoint we can find and try to reevaluate our overall belief set from there. Perhaps more importantly, even deciding what counts as the most epistemologically promising standpoint will be a deeply situated and empirical question. Keeping these difficult empirical problems in mind, however, the division of the mind into *Type-1* and *Type-2 processes* suggests some Spinozan solutions to the ethical problems that are generated by the automaticity of thought.

In the last chapter of *Ethics*, Spinoza suggests three ways in which reflective strategies can be mobilized in order to lessen the pernicious impact of *Type-1 processes*. His insight is simple: when the output of two mental processes conflict, one process must be modified. While we typically modify our conscious beliefs to bring them into line with the beliefs that have been produced by *Type-1 processes*, this change can occur in the other direction. *Type-2 processes* can be used to subvert automatic beliefs. However, in order for this to be the case we have to first understand the conditions that yield insidious stereotype-based judgments; once we do this, we can mobilize deliberate and reflective strategies in order to navigate our environment in a way allows us to avoid utilizing the *Type-1 processes* that we reflectively recognize as problematic. Following Spinoza, I suggest that there are three strategies for using *Type-2 processes* to minimize the pernicious impact of some *Type-1 processes*: 1) associating less problematic beliefs with a stimulus; 2) quickly exorcizing problematic beliefs; and 3) modifying our world to prevent problematic beliefs from being generated. In the remainder of this paper, I take each of these proposals in turn.

Let me begin with the strategy of using *Type-2 processes* to associate less problematic mental states with the stimuli that typically produce stereotypical judgments. On first blush, this may seem to be nothing more than the strategy of developing egalitarian attitudes; however, as I noted above, this strategy does not seem to be as effective as we might have hoped. Despite the fact that many of us recognize that the many of the stereotype-

based judgments we make are grounded on a distorted and prejudiced interpretation of the world, merely noting this is not sufficient to overcome years of behaving and judging on the basis of those stereotypes. As Spinoza (*Ethics* IVP2S) puts the point, "imaginations do not disappear through the presence of the true insofar as it is true, but because there are others, stronger than them, which exclude the present existence of things we imagine." Thus, the strategy of utilizing a *Type-2 process* to override a problematic *Type-1 process* requires a sensitivity to the conditions under which that *Type-1 process* operates, the cognitive distortions that it generates, and the affective components that make the outputs of that *Type-1 process* seem as plausible as they do. Once we have this in hand, it is possible to develop a strategy for modifying the *Type-1 processes* that we have come to recognize as problematic.

The sorts of strategies that will be successful have much in common with the strategies that have been developed as cognitive behavioral therapy (hereafter, CBT). There are various models of CBT; however, these models share a goal of replacing *Type-1 responses* that have been recognized as inappropriate with new responses that cancel the affective force of the initial judgment. For example, a person who is inclined to feel a deep and *apparently* insurmountable fear of academic failure can be trained to replace that fear with an alternative representation by focusing conscious thoughts on past successes. Although this strategy is difficult to enact, it is quite successful when it is structured around utilizing other thoughts that are emotionally powerful and inconsistent with the thought that must be avoided. Analogously, where *Type-1 processes* yield insidious stereotype-based judgments, Spinoza recommends recognizing the conditions under which a problematic judgment will be made and making incongruent judgments *in these cases.*[21] Thus, a person who knows that he will be likely to rely on irrelevant gender-based stereotypes in making a hiring decision can learn to consciously focus his thoughts on counter-stereotypical representations (e.g., by focusing on a female colleague in his field whose work he respects and who he takes to be a role-model). This is why being fragmented "is a good idea if you're going to have (or if you're stuck with) Spinozan belief-forming mechanisms that aren't perfectly reliable" (Egan forthcoming, 22). Because our minds consist of both *Type-1* and *Type-2 processes,* not all of our behaviors need to be directed by the systems that are generating the problematic beliefs.[22] We can, in some cases, utilize reflective *Type-2 processes* to inhibit our implicit biases so long as we have a good idea about the conditions under which these biases are likely to be activated and the conditions under which our behavior is likely to be driven by these biases.

However, in numerous cases, this CBT-based strategy of associating less problematic beliefs with a stimulus will be ineffective in overriding the *Type-1 processes* that produce stereotype-based judgments. In these cases, our knowledge of the conditions under which a stereotypical belief is likely to be produced allows us to recognize that a *Type-1 process* that yields confused and obscure ideas has produced this belief. When this occurs, a second Spinozan strategy of immediately exorcising problematic beliefs can be engaged in order to obviate the effects of the insidious stereotype-based judgment. This strategy, as with the first, falls within the scope of developing preferable strategies for social interaction by way of CBT. Consider once again the person who knows that she will become depressed in the face of perceived academic failure. Such a person can be trained to recognize the conditions under which she will judge that she has failed, and she can learn to immediately evaluate this judgment, to recognize that it is inconsistent with her other beliefs, and to thereby to reject this belief as false.

Again, this strategy is difficult to enact, and it requires a thoroughgoing understanding of the conditions under which problematic responses are generated. More importantly, in order to excise problematic beliefs, a person will already have to be convinced of the falsity of the outputs of this *Type-1 process*. There are, however, numerous cases where this technique has proved successful for people who struggle with depression or fear of failure; on this basis, we can hope that some *Type-2 processes* can be motivated to excise problematic stereotype-based judgments before they are incorporated into our overall cognitive economy. If we learn to recognize the conditions under which we make problematic stereotypical judgments, and we can learn to recognize them as problematic, then we can engage *Type-2 processes* in order to re-label these beliefs as false and then to excise them.

These two strategies, of course, paint an *overly optimistic* picture of our capacity to override stereotype-based judgments. In an ideal world where we were not constantly being exposed to stereotypical representations through both the overt ideological mechanisms of the mainstream media and through more subtle cues for categorization, we may be able to utilize strategies such as these. However, in the propaganda-filled world in which we live, these strategies are unlikely to be sufficient to overcome all of the insidious stereotypes upon which we rely in navigating our social world. Stereotypes are deployed to categorize in situations where a decision must be made quickly; and this means that we will tend to rely more heavily on stereotypical judgments under conditions of increased cognitive load. The data here are quite complicated. However, as Gilbert and Hixon (1991) have shown, although an increase in cognitive load does not increase the likelihood that a stereotype will be activated, it does increase the likelihood that a stereotype that has already been activated will be utilized in making later judgments. The worry, then, is that in conditions where a *Type-1 process* has already output a stereotypical representation, and where there is

[21] In fact, just as there is good evidence that repeated exposure can more deeply entrench stereotypical judgments, training by repeated exposure, focusing on the non-stereotypical properties of a stereotype-target, can influence the later judgments that will be made automatically by replacing the automatic stereotype with a less pernicious representation (cf., Kawakami et. al. 2000).

[22] Of course, how successful we will be depends "on which fragments have which functional roles in guiding behavior, and in how likely it is that you'll eventually subject the fragment containing the deliverances of perception to a process of evaluation where it won't pass muster." (Egan forthcoming, 23)

an increased cognitive load, we will be more likely to make subsequent judgments on the basis of the existing stereotypical representations. Thus, if the stereotype can be inhibited or eliminated by utilizing either of the first two strategies, cognitive load is less likely to generate problems. However, in cases where we fail to immediately excise or inhibit a stereotype, cognitive load will increase the likelihood that behaviors will be generated that accord with that stereotypical judgment. Moreover, as McRae and Bodenhausen (2000, 105) note, "an expansive literature has confirmed that category application is likely to occur when a perceiver lacks the motivation, time, or cognitive capacity to think deeply (and accurately) about others". Given that this is the case, even attempting to purge stereotype-based judgments might be a strategy that offers too little, too late.

This brings me to the final, most effective option advanced by Spinoza. Once we have developed an adequate account of the causal relations between problematic *Type-1 processes* and the stimuli that trigger them, we can begin to develop strategies that allow us to avoid the situations in which insidious stereotype-based representations are produced. At this point, however, it becomes clear that the only way in which we will be able to adequately modify our psychology is by modifying the world in which we live. This strategy has both local and global variants, which I take in turn.

The local strategy is nicely developed in response to the CV study described in Section 1.[23] Steinpreis and her colleagues (1999) found that academic psychologists were more likely to recommend hiring a male than a female, more likely to offer positive evaluations of a male's contributions to the field, and, more likely to want evidence that a female had done the work shown on her CV on her own. These findings suggest that women tend to receive less benefit from their achievements than men do. Fortunately, there are fairly effective strategies for preventing this result. However, in order to address this issue, the conditions under which CVs will be evaluated must be modified in order to mitigate the impact of gender on deciding whom to hire. In an attempt to develop strategies that can achieve this goal, the STRIDE (Strategies and Tactics for Recruiting to Improve Diversity and Excellence) committee at the University of Michigan has developed an incredibly useful set of tools for evaluating potential job candidates.[24] Instead of relying on global evaluations of the way that interactions with a candidate 'feel', the STRIDE program recommends that a specific strategy be developed that focuses on the particular qualities that are being looked for in a candidate. For example, all candidates are evaluated on a scale ranging from excellent to poor (with the alternative possibility of "unable to judge") for such things as potential for research productivity, fit with the departments specified goals, and potential for attracting and advising students.

These strategies may seem too simple to be of use. However, with SBF in mind, it becomes clear precisely why these strategies are effective. By focusing on particular sorts of facts about a candidate, we are able to provide orienting information that allows us to modify the background categories that are used for interpreting a person's qualifications. Thus, rather than relying on the immediate classification of a person *as female*, for example, focusing on this sort of information instead activates the category *scholar* or *potential colleague*. While we may be inclined to draw the unwarranted inference that *a woman* will be less committed to research because she will also want to be a mother, we are significantly less likely to consider the effects of parenthood on scholarship as such. As Shelly Correll and Stephan Benard (2007) have shown, although we intuitively judge that mothers are less competent and less committed to research, we do not make the same judgment about fathers. By making the gender-based category less salient, the information is more likely to be processed without activating gender-based stereotypes for classifying the individual.

The point, here, is that although we can recognize merit when merit is what we are paying attention to, when gender-based stereotypes are active this can focus attention on factors that are irrelevant to the interpretation of the data with which we have been presented. By carefully specifying the sorts of considerations that are significant for an evaluation, and by collecting only information that focuses explicitly on these considerations, we leave ourselves with less room to rely on the results of *Type-1 processes* that might deliver problematic stereotype-based judgments. I cannot address all of the important strategies that have been developed by the STRIDE committee, as well as various ADVANCE programs, for improving the climate for women in academia. However, it would serve everyone in academic positions well to take the time to investigate these strategies more thoroughly. Unfortunately, however, I remain unconvinced that even these strategies can guarantee that we will always avoid making irrelevant stereotype-based judgments, and this brings me to my final suggestion, the global strategy for modifying the world in which we live.

At the end of the day, there will always be a substantial worry that these sorts of changes will be insufficient to rid us of all of our problematic stereotypical generalizations. As developmental psychologists such as Lev Vygotsky (1934/2000) and Jerome Bruner (1975) have argued, many of our representations of the world are scaffolded onto the representations of others. That is, our capacity to understand both 'self' and 'other' arise, at least in part, through our early interactions with other people in our community. Because of the ways in which others categorize, and because of the way that our language causes us to categorize, we develop strategies for carving up the world that may not suit its joints. This is not, of course, to claim that we have no capacities to represent the world apart from those that we acquire through culture; rather, the claim is that our categorization of social phenomena is intimately tied to the social structures in which we learn to categorize. While it may indeed be possible to retool our conceptual repertoire later in life, it is

[23] The local strategies discussed in what follows were developed on the basis of Virginia Valian's online "Tutorials for Change" and the resources available online through the Michigan ADVANCE and STRIDE programs (see the Works Cited section for complete citations),

[24] Many thanks are due to Alison Wylie for pointing me to this program.

unclear to what degree this is possible. Perhaps, the Spinozan would argue, the only way to rid ourselves of the insidious stereotypes that generate, propagate, and facilitate the asymmetric power relations that pervade our world is through a change in the conditions for the acquisition of schemas for categorizing the things that we find in the world. Perhaps, that is, eliminating the pernicious effects of *Type-1 processes* that produce stereotype-based judgments requires eliminating the social structures that allow us to acquire pernicious gender-based stereotypes in the first place. If this is correct, then the Spinozan ideas that have often been seen as offering a radical critique of existing social organizations will be the only viable option for overcoming our insidious stereotype-based judgments!

6. WORKS CITED:

Aarts, H., and A. Dijksterhuis (2003). "The silence of the library." Journal of personality and social psychology **84**.

Bargh, J., M. Chen, and L. Burrows (1996). "Automaticity of social behavior." Journal of personality and social psychology **71**: 230-244.

Bertrand, M., and S. Mullainathan (2003). "Are Emily and Greg more employable than Lakisha and Jamal?" Working paper series; MIT department of Economics.

Bruner, J. (1975). Language as an instrument of thought. Problems of language and learning. A. Davies. London, Heinemann.

Bulthoff, I., and F. Newell (2004). "Categorical perception of sex occurs in familiar but not unfamiliar faces." Visual cognition **11**(7): 823-55.

Carruthers, P. (2007). "The illusion of conscious will." Synthese **159**(2): 197-213.

Chaiken, S., and Y. Trope (1999). Dual-process theories in social psychology. New York.

Corell, S., S. Benard, and I. Paik (2007). "Getting a job: is there a motherhood penalty." American journal of sociology **112**: 1297-1338.

Cutting, J., and E. Cafarelli (1977). "Recognizing friends by their walk." Bulletin of psychonomic society **9**: 253-356.

Descartes, R. (1988). The philosophical writings of Descartes. Cambridge, Cambridge University Press.

Dijksterhuis, A., and A. Van Knippenberg (1998). "The relationship between perception and behavior, or how to win a game of trivial pursuit." Journal of personality and social psychology **74**: 865-77.

Dostoyevsky, F. (1863/1997). Winter notes on summer impressions. Evanston, IL, Northwestern University Press.

Egan, A. (forthcoming). "Seeing and believing." Philosophical studies.

Faucher, L., E. Machery, and D. Kelly (in preparation). "The 'psychological pluralism' of racial prejudice."

Festinger, L., H. Riecken, and S. Schachter (1956). When prophecy fails. Minneapolis, University of Minnesota Press.

Fiske, S., and S. Taylor (1991). Social cognition. New York, McGraw Hill.

Gilbert, D. (1991). "How metal systems believe." American psychologist **46**: 107-19.

Gilbert, D. (1993). The assent of man. The handbook of mental control. D. W. a. J. Pennebaker. White Cliffs, NJ, Prentice Hall**:** 57-87.

Gilbert, D., and J.G. Hixon (1991). "The trouble of thinking: Activation and application of stereotypic beliefs." Journal of personality and social psychology **60**(4): 509-517.

Gilbert, D., D. Krull, and P. Malone (1990). "Unbelieving the unbelievable." Journal of personality and social psychology **59**: 601-613.

Gilbert, D., R. Tafarodi, and P. Malone (1993). "You can't not believe everything that you read." Journal of personality and social psychology **65**: 221-233.

Griffiths, P., and A. Scarantino (2005). Emotions in the wild. The Cambridge handbook of situated cognition. P. Robbins, and M. Aydede. Cambridge, Cambridge University Press.

Johnson, K., and L. Tassinary (2005). "Perceiving sex directly and indirectly." Psychological science **16**(11): 890-897.

Kawakami, K., J. Dovidio, J. Moll, S. Hermsen, and A. Russin (2000). "Just say no (to stereotyping)." Journal of personality and social psychology **78**(5): 871-88.

Kay, A., S. Wheler, J. Bargh, and L. Ross (2004). "Material Priming." Organizational behavior and human decision making **95**: 83-96.

Ko, S., D. Muller, C. Judd, and D. Stapel (in press). "Sneaking in through the back door." Journal of experimental social psychology.

Larsen, K., M. Reed, and S. Hoffman (1980). "Attitudes of heterosexuals toward homosexuality." Journal of sex research **16**: 245-257.

Lewin, A. and L. Duchan (1971). "Women in Academia." Science 173: 892-895.

MacRae, C., and G. Bodenhausen (2000). "Social cognition: Thinking categorically about others." Annual review of psychology **51**: 93-120.

MacRae, C., G. Bodenhausen, A. Milne, and J. Jetten (1994). "Out of mind but back in sight." Journal of personality and social psychology **67**: 808-17.

Mandelbaum, E. (in preparation). The architecture of learning. Philosophy, UNC - Chapel Hill.

Monteith, M., C. Spicer, and G. Tooman (1998). "Consequences of stereotype suppression." Journal of experimental social psychology **34**(4): 355-77.

Schaller, M., and S. Neuberg (in press). Intergroup prejudice and intergroup conflicts. Foundations of evolutionary psychology. C. C. a. D. Krebs. Mahwah, NJ, Lawrence Erlbaum Associates**:** 399-412.

Spinoza, B. (1677/1991). Ethics. Indianapolis, IN, Hackett.

Steele, C., and J. Aronson (1995). "Stereotype threat and the intellectual test performance of African-Americans." Journal of personality and social psychology **69**(5): 797-811.

Steinpreis, R., K. Anders, and D. Ritzk (1999). "The impact of gender on the review of curricula vitae of job applicants and tenure candidates." Sex roles **41**: 509-528.
Unknown. "ADVANCE Program." Retrieved 20 April, 2008, from http://sitemaker.umich.edu/advance/home.
Unknown. "STRIDE Committee." Retrieved 20 April, 2008, from http://sitemaker.umich.edu/advance/stride.
Valian, V. (2005). Why so slow? Cambridge, MIT Press.
Vygotsky, L. (1934/2000). Thought and language. Cambridge, MIT PRESS.
Wegner, D. (1994). "Ironic processes of mental control." Psychological review **101**: 34-52.
Wegner, D., and J. Bargh (1998). Control and automaticity in social life. Handbook of social psychology. S. F. D. Gilbert, and G. Lindzey. New York, McGraw Hill**:** 446-96.
Wegner, D., D. Schneider, S. Carter, and T. White (1987). "Paradoxical effects of thought suppression." Journal of personality and social psychology **53**: 5-13.
Wegner, D., R. Erber, and P. Raymond (1991). "Transactive memory in close relationships." Journal of personality and social psychology **61**: 923-29.