

THE VIEW FROM NOWHERE THROUGH A DISTORTED LENS:

THE EVOLUTION OF COGNITIVE BIASES FAVORING BELIEF IN FREE WILL

By Kip Werking

[pdf [version](#); htm [version](#); doc [version](#)]

“In the distant future I see open fields for far more important researches. Psychology will be based on a new foundation, that of the necessary acquirement of each mental power and capacity by gradation. Light will be thrown on the origin of man and his history.”
—Charles Darwin (1859: 449)

“One must view a wicked man, like a sickly one—We cannot help loathing a diseased offensive object, so we view wickedness.—it would however be more proper to pity than to hate & be disgusted.”
—Charles Darwin (1987: 608)

I. INTRODUCTION

Although philosophers have confronted the evidence for human irrationality in other contexts (Mele 1987; Stein 1996), there is one area in which this evidence may be of great relevance but has yet to be discussed—the free will problem. Instead, a vast ocean seems to separate the respective literatures on free will and human irrationality. One might say that there is a lacuna in the literature about the lacunas in our minds. The result is that philosophers who study free will work in a context which tends to presuppose that humans are rational.

The silence about human irrationality is related to another silence about the demanding concepts many have for free will. It is telling that compatibilists about determinism and free will defend concepts of that power which nobody would deny humans can satisfy. Their arguments against libertarian views are often limited to show that such concepts are inconsistent with actual human agency and not that the stronger notion of libertarian freedom does not govern the relevant terms in the debate. As against the non-realist, this option is unavailable to the compatibilist and so compatibilists often

give non-realist views on free will little attention (Strawson 1994) or grant that they are defending something different than the traditional notion of free will (Dennett 2003). Indeed, one prominent compatibilist has characterized the stronger notion of free will as “a sort of metaphysical megalomania” (Fischer 2006b). Compatibilists are just now beginning to research the issue of whether their conception of free will is too weak and the initial results are both mixed and intriguing (Nahmias et al. 2004, forthcoming; Nichols 2004, forthcoming). Given the presumption of human rationality, compatibilists may assume that more demanding concepts of free will are irrational and so pass over them in silence.

In this article, I attempt to show, by bridging the ocean between these two literatures and rebutting the presumption of human rationality, that libertarian and non-realist concepts of free will are not so unreasonable. First, this introduction (I) will distinguish between two kinds of control and show how this distinction can frame the free will problem. The introduction will also describe the ways in which cognitive biases may evolve. In part two (II), I analyze the psychology of human decision making. In particular, I focus upon certain critical decisions and argue that a rational agent would look upon these options as one would look upon Thomas Nagel’s View From Nowhere. But certain cognitive biases allow people to view this View From Nowhere through a distorted lens—and therefore to avoid its paralyzing sting. I argue that these biases evolved because they are adaptive in various ways. In part three (III), I will show how certain cognitive biases lead people to feel they have more control and responsibility, with respect to their decisions, than they actually do. I speculate that these biases evolved because they falsely signal certain attractive traits to others. In part four (IV), I

will show how one profound cognitive bias, the just world phenomenon, and several others related to it, would further incline one to believe in free will. I speculate that these biases may have evolved because they were adaptive or evolved despite being non-adaptive. I conclude, in part five (V), that a multitude of cognitive biases may infect human thinking about free will. The vulnerabilities in the human psyche suggest that libertarian and non-realist concepts of free will may capture, more than their compatibilist alternatives, the usage which governs the meaning of terms in the debate about free will.

First, it is important to distinguish, in the ancient controversy between those who defend the existence of free will and those who assail it, between two kinds of *control*.ⁱ The first kind of control is *actual* control. This is the sort of control human beings have over their lives. It is such that, *given* one's ultimate ends and purposes, one can make choices according to those ends and purposes and act accordingly.ⁱⁱ

But there is another kind of control. To understand this second type of control, first consider what I will call *novelist* control. This is the sort of control that novelists have over their characters. According to novelist control, an author may choose not just how a character makes decisions in accordance with certain ultimate ends and purposes, but also choose what these ends and purposes will be. The ultimate ends and purposes are themselves *chosen*—not *given*. In this way, the *scope* of novelist control is much wider than that of actual control. A novelist has much more control over a character's life trajectory than the character does. For this reason, some accuse actual control of being *shallow* (Smilansky 2003).

There is a superficial sense in which novelist control is buck-stopping (Dennett 2003: 99). In the ordinary case, to tell a story about some agent's life, one would need to say something about how that agent's ultimate ends and purposes arose. For example, one might want to tell a story about how the agent evolved, or inherited certain genetics, or lived in a certain childhood environment. Even this may not be entirely satisfying because one might always inquire: "why that evolutionary path, why those genetics, why that childhood environment?" At some point, simple logic shows that these questions must go unanswered. There is a temptation, however, to feel that, in the novelist scenario, the novelist provides a robust and sufficient explanation for agent's life. The novelist is responsible for the meticulous creation of every detail in the character's life. Surely this author can provide a sufficient explanation for why the agent's life story followed the path that it did. One might ask why the character's life unfolded as it did and the author could say "I wanted to create a character with property X and who lives a life with features Y and Z." But the question remains: "why did you desire to create a character with *that* property and who lives a life with *those* features?" Again, such questions must eventually go unanswered.

When one considers novelist control in the context of this ancient controversy, however, one encounters a tantalizing possibility—to what extent might agents have novelist control over their *own* lives? I will call this sort of control *novelist* control*. If people have novelist* control over their own lives, then they make decisions not just in accordance with their ultimate ends and purposes, but they furthermore determine what these ultimate ends and purposes will be. Actual novelists already approach this level of control over their characters. To the extent that their stories lack in the richness of detail

present in reality, one can imagine how God or an advanced being (perhaps a radically enhanced human being in a future civilization) might design the lives of actual humans and not just characters in novels. Greene and Cohen consider just such a scenario:

“It is very simple, really. I designed him. I carefully selected every gene in his body and carefully scripted every significant event in his life so that he would become precisely what he is today. I selected his mother knowing that she would let him cry for hours and hours before picking him up. I carefully selected each of his relatives, teachers, friends, enemies, etc. and told them exactly what to say to him and how to treat him. Things generally went as planned, but not always. For example, the angry letters written to his dead father were not supposed to appear until he was fourteen, but by the end of his thirteenth year he had already written four of them. In retrospect I think this was because of a handful of substitutions I made to his eighth chromosome. At any rate, my plans for him succeeded, as they have for 95% of the people I’ve designed. I assure you that the accused deserves none of the credit.” (Greene and Cohen 2004)

There is one crucial point to understand about this second kind of control. The point is that the apparent similarity between novelist control and novelist* control is *deceptive*. Of course, novelist* control is logically impossible. The act of choosing one’s character presupposes the existence of a character according to which one can make these choices. Character cannot create itself in a vacuum. But novelist control, unlike novelist control, *is* possible.

So the apparent similarity between novelist control and novelist* control is deceptive. To the extent that novelist control and novelist* control *appear* similar, however, one can understand how people might make this mistake. It is not *obvious*, for example, that one might have control over the ultimate ends and purposes of another but no such control over one’s own ultimate ends and purposes. As the rest of this article will explain, certain cognitive biases might enable this illusion of feeling that one has

more than actual control over one's lives—and that one approaches having novelist* control.

This distinction between actual control and novelist control helps *frame* the ancient dispute between those who defend the existence of free will and those who assail it. First, one might characterize the dispute as one just about the powers humans have. On this view, both parties agree about the relevant concept of free will and then examine human nature to see whether human being can, in fact, satisfy this concept. Indeed, the distinction between actual and novelist* control suggests that all parties might agree that the relevant notion of control being tested is quite strong—not unlike novelist* control. One happy consequence of this characterization is that it is conducive to empirical investigation. Scientists can examine human nature and see whether human beings have such powers. The dispute between libertarians and non-realists has this character (Pereboom 2001: 69-88; Nichols, forthcoming).

Second, one might characterize the dispute between these two parties as one about what the relevant concept of free will is. So according to this characterization, both parties would hold their understanding of human nature constant and then consider whether this notion of human nature satisfies the relevant concept of free will. The distinction between actual and novelist* control suggests that all parties agree that the account of human agency is relatively weak—not unlike actual control. One must then determine whether this account of human agency can satisfy the relevant concept of free will. Unfortunately, this view is much less conducive to empirical investigation. It is more difficult to study the exact meaning, if any, of the terms “free will” and “moral responsibility” than it is to study whether indeterministic processes in the brain enhance

the control humans have over their decisions.ⁱⁱⁱ The dispute between compatibilists and non-realists has this character.

The dispute, according to first characterization, is easier to solve than the dispute, according to the second characterization, because human beings are just more available to inspection than the terms “free will” and “moral responsibility” are. Similarly, the human genome is more exact than the concept of “love” and the Mona Lisa is more definite than the concept of God. That those who defend the existence of free will may engage in both tacit and overt acts of *revisionism* only exacerbates this difficulty. Whether indeterministic processes in the brain enhance the control we have over our decisions will never change; on at least some views, however, the terms of “free will” and “moral responsibility” might have meant one thing yesterday but should mean something different today (Vargas 2005).

Perhaps despairing at the prospects of winning the first battle, the majority of philosophers who defend the existence of free will have staked their claims in the murkier territory of the second battle. Considering the difficulty of resolving the dispute, according to this second characterization, the prospect of progress in the near future is grim. In contrast, I suggest a third way of approaching the dispute which promises imminent progress. Whereas the first and second characterizations of this dispute contrast the powers people have with the powers required by free will or moral responsibility, this third strategy contrasts both the powers and responsibilities that people have with the powers and responsibilities people *think* they have.

The answer to this third question may be relevant to the ancient dispute for two reasons. First, whereas libertarians can deny that people lack libertarian free will, they

might not be able to deny, in the face of contrary evidence from the cognitive sciences, that people think they have more power than they actually do—and this illusory power bears a resemblance to libertarian free will. Second, whereas compatibilists can deny that the terms “free will” and “moral responsibility” refer to concepts demanding more than actual control, they might not be able to deny, in the face of contrary evidence from the cognitive sciences, that ordinary responsibility practices presuppose an inflated sense of control and responsibility. In particular, this third strategy avoids any danger of *revisionism*. Clever philosophers in ivory towers might settle for a weaker notion of freedom and responsibility than the ordinary use of these terms (if any) would demand; the folk or non-specialists in the labs of cognitive scientists do not have this luxury.

In this way, progress according to this third strategy might help to rebut the presumption of human rationality. Allegedly rational belief in free will has been the traditional default; the burden of proof was on those who would reject this orthodox view of human agency. For example, some suggest that the failure of the non-realist view to satisfy the standards of conservatism (whereby folk intuitions survive philosophical inspection) explains its unpopularity:

“But this view nowadays has few unabashed subscribers, and its proponents through the history of philosophy have been relatively few in number. Indeed, it is tempting to suppose that hard determinism is the exception that proves the rule, since a likely diagnoses of its unpopularity is its failing standards of conservatism; as Vargas (2005: 401) remarks, approaches commending substantial revision of folk morality have typically faced a ‘broad and ready skepticism.’” (Dorris et al., forthcoming)

Indeed, without a decisive argument or experiment, those who defend the existence of free will and moral responsibility might have difficulty even understanding what motivates the skeptic. If the defender of non-realism about free will can

demonstrate, however, that people engage in widespread and systematic irrationality when making attributions of moral responsibility and estimations of the control they and others have over their lives, then those who defend orthodoxy can understand what motivates the non-realist's view. The non-realist has just identified certain cognitive biases and revised certain beliefs in accordance with this new information. This would help rebut the presumption of human rationality and the burden might shift to the defender of orthodoxy to show how such seemingly irrational beliefs can survive the discovery of these biases.

There are two ways to proceed along this front. First, philosophers might investigate folk beliefs in the areas that interest them. Limited funds and scientific training may result in polls that are lacking in rigor and sample size. This seems to be the strategy of the experimental philosophy movement. That movement has its virtues but it is too small and too new to promise much progress in the near future.

An alternative strategy is to analyze data which cognitive scientists have *already* collected. One might suspect that the different motivations of the philosopher and the cognitive scientist would cause any such data to be of limited relevance. This article will argue, however, that the field of *cognitive biases and heuristics* provides an abundance of data which can help settle the dispute between those who defend the existence of free will and those who do not.

In exploring the dispute according to the third characterization, I have followed this second strategy of investigating the relevant cognitive science literature. To my surprise, I discovered at least *fifteen* cognitive biases favoring belief in free will and *none* favoring disbelief in free will. As one prominent scholar in this area has noted,

“[n]umerous psychological studies on blame and responsibility, as well as the perplexing outcomes of high-profile criminal and civil trials, demonstrate that everyday blamers are capable of violating virtually every rational prescription that moral philosophers, legal scholars, and rational decision theorists hold dear” (Alicke, forthcoming).

Although just showing how certain cognitive biases relate to the free will problem would be sufficient to create a noteworthy paper, this article will go further and explain how these biases might have *evolved*. These remarks will, like those found in most evolutionary psychology, be necessarily *speculative*. Indeed, this approach puts the conclusions of this article in double jeopardy because both the heuristics and biases literature (Gigerenzer 1996) and the evolutionary psychology literature (Rose & Rose 2000; Buller 2005) have their critics—including each other (Haselton & Buss 2003). Despite the lingering controversy over these subjects, however, this article will suppose that both literatures have their several virtues. Furthermore, the remarks on the evolutionary origins of these cognitive biases will provide a *context* for the philosophical content in this article and further the use of *science* to solve philosophical problems. Of course, not all of these speculations will be accurate. Furthermore, not all cognitive biases will invite just one evolutionary explanation; any one bias might suggest a variety of explanations and these may be either adaptive or non-adaptive. Nevertheless, exploring the evolutionary origins of belief in free will may strengthen the article’s larger argument.

Although there are several possible explanations for cognitive biases, this article will focus upon *evolutionary* explanations. Not all possible explanations are evolutionary: cognitive biases may represent, for example, systematic performance errors

(Stein 1996: 8-9) or the failure of participants in an experiment to properly construe the given task (Gigerenzer 1996). Furthermore, not all evolutionary explanations are adaptive—natural selection is not omnipotent. In explaining any given trait, one must remember the importance of concepts such as pleiotropy (Williams 1957), exaptations (Gould & Lewontin 1979), and genetic drift (Wright 1932).

But many evolutionary explanations are adaptive. Cognitive biases may represent vestigial traits that, although they were adaptive in the earlier times, are no longer so (Haselton & Buss 2003). Furthermore, traits may be adaptive in a variety of ways. First, they may contribute to one or both of the individual's longevity (natural selection in the classic sense; Darwin 1859) or reproductive success (*sexual*, as distinct from natural selection; Darwin 1871: 256). There is also a controversy, amongst those who research the evolution of cognitive biases, as to how these biases evolved. The heuristics and biases literature suggests that these biases represent adaptive heuristics which compensate in computational or metabolic efficiency for what they lack in accuracy (Tversky & Kahneman 1974). But a growing body of literature in evolutionary psychology suggests that such biases are adaptive, regardless of computational limitations, because humans evolved to maximize fitness and not epistemic accuracy; so, to the extent these two goals conflict, fitness will prevail at the expense of truth (Haselton & Buss 2003). In particular, Buss and Haseltine attempt to explain the evolution of cognitive biases according to Error Management Theory (EMT) (2000, 2003). EMT predicts that such biases will evolve when the following conditions obtain: "(1) when decision making poses a significant signal detection problem (i.e., when there is uncertainty); (2) when the solution to the decision-making problem had recurrent effects on fitness over evolutionary history; and

(3) when the aggregate costs or benefits of each of the two possible errors or correct inferences were asymmetrical in their fitness consequences over evolutionary history” (Haselton & Buss 2003: 31).

The cognitive biases and heuristics approach and EMT may not necessarily conflict. One can understand this potential harmony in the context of Buss and Haselton’s notion of a smoke alarm as an example of error management logic (2003: 31). Smoke alarms are intended to be hypersensitive because the dangers of a false negative are so much greater than the dangers of a false positive. The high ratio of false positives to false negatives is an intended feature of their design; by analogy, the human terror of harmless snakes may be a feature of our design. But fire alarms engage in this error management *because* they have limited computational resources. A fire alarm with a supercomputer and sophisticated spectrometer could afford to make less false positives by also avoiding false negatives—even if the ratio of safe errors to dangerous errors remains biased towards safe errors. In this way, the cognitive biases and heuristics approach and EMT may complement each other.

To illustrate the EMT approach, consider what is, to my knowledge, the only evolutionary explanation ever proposed for belief in free will: “The Illusion of Free Will Evolves” by Tamler Sommers (forthcoming). Sommers’ approach is also unique to the extent that it proposes that cognitive enhancement aggravated, rather than ameliorated, the error management solution to adaptive problems. He describes how “cognitively sophisticated” (CS) creatures might begin to doubt the rationality of their reactive attitudes. For example, a CS creature might doubt that enacting retribution against a rival is worth the risk of harm to one’s self. Sommers’ insight is the suggestion that humans

evolved belief in free will and retribution to *compensate* for the pacifying affect of this cognitive enhancement.

Like Sommers, I suggest a dual process explanation for belief in free will. Nichols has suggested that a dual process theory explains intuitions about free will (forthcoming) and Greene has defended a similar dual process explanation for deontology (forthcoming). Such theories are consistent with the seminal work by Haidt showing that people make quick judgments about morality and only later justify these judgments through a process of rationalization (2001). Dual process theories all suggest that our evolved psychologies contain a number of domain specific circuits. For example, Steven Pinker argues that:

“[T]he mind is organized into cognitive systems specialized for reasoning about object, space, numbers, living things, and other minds; that we are equipped with emotions triggered by other people (sympathy, guilt, anger, gratitude) and by the physical world (fear, disgust, awe); that we have different ways for thinking and feeling about people in different kinds of relationships to us (parents, siblings, other kin, friends, spouses, lovers, allies, rivals, enemies); and several peripheral drivers for communicating with others (language, gesture, facial expression).”^{iv}

Such domain specific neural circuitry may provide fast and cheap calculation whereas more general processing in the cortex may provide slower and more expensive, but more thorough and accurate, computation. These circuits may account for both the reactive attitudes people feel towards others as well as the beliefs people maintain about human agency. Because they are faster and cheaper, but less accurate, than more general processing in the cortex, the two processes may come into conflict—reflection might make the illusion of free will disappear. This is consistent with the remarks of Watson (1986), Strawson (1986), and Nagel (1986) that, although reflection can incline one to doubt the existence of free will, this skepticism is difficult to maintain. As soon as one

exits calm reflection, the faster and cheaper, but inaccurate, domain specific circuits begin to dominate again.

Unlike Sommers' view, however, this article will not argue for any single adaptation explaining belief in free will. Instead, this article will provide an *indirect* argument by considering a multitude of cognitive biases which may combine to explain, in part, belief in free will.

This introduction has noted a silence about human irrationality in the literature on free will. To address this silence, it first distinguished between two kinds of control and showed how these kinds of control frame the free will debate. Second, it suggested that the abundant literature on cognitive biases and heuristics will help the debate over free will progress. Finally, the introduction discussed two evolutionary explanations for cognitive biases (as heuristics and as error management) and suggested its own overarching explanation for belief in free will. In section two (II), the article will begin by considering several cognitive biases which may favor belief in free will in light of one philosopher's notable work on the subject.

II. THE VIEW FROM NOWHERE THROUGH A DISTORTED LENS

Less than two decades ago, Thomas Nagel (1986) and Galen Strawson (1986) suggested not just that free will does not exist but that it is also *impossible*.^v In particular, Nagel distinguished between the *subjective* viewpoint, which houses our precious beliefs, and the *objective* viewpoint, which threatens to destroy them. The subjective viewpoint presents metaphysical alternatives before agents without inquiring into the forces that shape their characters. It presupposes, and demands, that people are the ultimate creators

and sustainers of their own ends and purposes. But reflection shows that nothing can satisfy this conception of human agency.

Nagel writes that the objective view “takes in not only the circumstances of action as they present themselves to the agent, but also the conditions and influences lying behind the action, including the complete nature of the agent himself.” So, in Nagel’s view, the objective view is problematic because it presupposes that the agent has a given character and that this character constrains its future choices. But Nagel also describes the objective view as presupposing no such character. In this sense, the objective view requires one, in order to be free, “to act from a standpoint completely outside ourselves, choosing everything about ourselves, including all our principles of choice—creating ourselves from nothing, so to speak.” This description suggests that the agent, in uneasy tension between its presently blank and future selves, must look out over the space of possible characters and choose, *ex nihilo*, what kind of agent to be. This last image deserves Nagel’s lovely title: the “View from Nowhere.”

The View from Nowhere has a paralyzing sting. As Nietzsche wrote, one just cannot “pull oneself up into existence by the hair, out of the swamps of nothingness” (1886: §21). By analogy, although one might have novelist control over another person, one cannot have novelist* control over one’s self. Yet people regard themselves as free and responsible. Furthermore, they seem to accomplish this by embracing an illusory feeling of control, and not by weakening their demands for such control. In the same paragraph, Nietzsche noted that “[t]he desire for ‘freedom of the will’ in the superlative metaphysical sense, [] still holds sway, unfortunately, in the minds of the half-educated.” This article will further describe the abundant evidence showing that people regard

themselves as possessing more freedom and responsibility, with respect to their lives, than they do in fact have. Somehow, people jump out of the View from Nowhere.

This article's central thesis is that people accomplish this feat by looking at the View from Nowhere through a *distorted* lens. In particular, certain cognitive biases warp the View from Nowhere such that it presents agents with reasons for preferring, from amongst the space of possible characters, one particular character—their later selves. These biases are the *endowment effect* and the *outcome effect*. Similar biases, such as *contamination* and the *mere exposure effect*, further this phenomenon. Moreover, the *fundamental attribution error*, working in concert with *anthropomorphic bias* and *asymmetric attributions of blame*, provides a ready vehicle in which the endowment effect and outcome bias may infect human thinking about freedom and responsibility. These biases render The View from Nowhere into the View from Somewhere.

First, to understand how the endowment effect might allow people to jump out of the View from Nowhere, one must appreciate that the free will problem is about *gifts*. In particular, the free will problem involves the difficulty of embracing the gift of one's character. For example, one prominent libertarian bases his position upon the existence of indeterministic “Self-Forming Actions” (Kane 1996). Similarly, one prominent defender of non-realism about free will has observed that:

“Compatibilists claim that this is the right thing to say. They believe that to have free will, to be a free agent, to be free in choice and action, is simply to be free from constraints of certain sorts. Freedom is a matter of not being physically or psychologically forced or compelled to do what one does. Your character, personality, preferences, and general motivational set may be entirely determined by events for which you are in no way responsible (by your genetic inheritance, upbringing, subsequent experience, and so on). But you do not have to be in control of any of these things in order to have compatibilist freedom. They do not constrain or compel you, because compatibilist freedom is just a matter of

being able to choose and act in the way one prefers or thinks best given how one is.” (Strawson 1998, 2004)

This character is *constraining* in the sense that, although it might later modify itself, it will only do so in accordance with its original makeup. In Saul Smilansky’s words, free will does not exist because people are just the “unfolding of the given.” We are not the unfolding of the chosen. The quality of human nature which remains disturbing is this *given-ness*. This given-ness threatens to alienate people from themselves because one might always question whether one is as one ought to be. Yet people somehow jump out of the View from Nowhere and embrace the gifts of their characters.

One subtle way in which people might accomplish this feat involves the *endowment effect*. This effect was discovered by Kahneman and Tversky (1979) and named by Thaler (1980). According to the endowment effect, ownership confers value upon goods and services. Although some have questioned the existence of the effect (Shogren et al. 1994; Plott & Zeiler 2005), a large number of experiments have confirmed it. Recently, Nayakankuppam and Mishra repeated the traditional experiment by situating eighty-eight students next to a coffee mug, telling them that they were either buyers or sellers, and then asking them for their price (2005). Nayakankuppam and Mishra reported that “The basic endowment effect emerged with sellers quoting significantly higher reservation prices (\$5.59) than buyers, (\$4.43), $F(1, 82) = 5.64$, $p < .02$.”

This endowment effect is similar to, and perhaps a subset of, the *mere exposure effect*. Psychologists first noticed this effect in the nineteenth century (Fechner 1876); in modern psychology, the most prominent psychologist to research it is Zajonc (1968).

According to the mere exposure effect, just the fact that one is familiar with something will increase one's liking it. A recent meta-analysis of research on the mere exposure effect concluded that "[t]he first 20 years of research on Zajonc's (1968) mere exposure effect leaves little doubt that the exposure–affect relationship is a robust, reliable phenomenon" (Bornstein 1989).

Several scholars have speculated on the evolutionary origins of the endowment effect. David Friedman suggests that an endowment effect reflects territoriality and notes that such territoriality has been observed in many species.^{vi} Friedman argues that such territoriality was adaptive in earlier times, even if it is no longer adaptive, because it motivated humans to “fight very hard” for their possessions. Others have observed the precise endowment effect in capuchin monkeys (Chen et al. 2006). This led these researchers to conclude, as Friedman did, that the bias is innate and not learned. From the perspective of economics, Steffen Huck, Georg Kirchsteiger and Jörg Oechssler have provided a precise mathematical model for how an endowment effect evolved because it “improved one's bargaining position in bilateral trades” (2005). Similarly, Herbert Gintis has developed a game theoretic model to explain how the endowment effect enabled the natural evolution of private property (Gintis, forthcoming). In the context of the gift of one's character, including one's genetic endowment, one must also note that natural selection could not favor any tendency to doubt or disown this character. Genes which make one irrationally prefer them will thrive at the expense of genes which allow one to approach them with skepticism.

Scholars have also speculated on the origins of the mere exposure effect. Bornstein notes that it is adaptive “for adults to prefer the familiar over the novel”

because, even if there are advantages to exploring new territory, “some risk is inherent in any venture into the unknown” (1989). In contrast, it is adaptive for infants to prefer the novel over the familiar because they need to grow familiar with their environment and can rely upon their mothers to protect them. Bornstein further notes that his suggested explanation is consistent with the findings that delay enhances, and boredom mitigates, the mere exposure effect. He proposes, therefore, that both the adult preference for familiarity and the infant preference for novelty have “evolved into the natural repertoire of human behavior.”

The endowment and mere exposure effects provide subtle but compelling explanations for how people jump out of the View from Nowhere. The View from Nowhere derives its sting from the vacuum of values existing before the agent’s birth. From the agent’s vantage point, it cannot yet choose which character to have because any such choice presupposes an existing character. So the agent remains paralyzed before this vista of character space because all possible characters are equally desired and equally abhorred. But the endowment effect and mere familiarity affect give people a reason for preferring one of these characters from amongst the others—their later selves. According to these effects, people will prefer what they own and that with which they are familiar. To the extent that people forget that, according to the View from Nowhere, their later selves can have no rational influence upon their choice, they will prefer the selves they have been given. Other cognitive biases show that people will tend to forget just this sort of fact.

One cognitive bias which would help people forget this sort of fact is *outcome bias*. According to this cognitive bias, people consider the eventual outcome of a

decision as a relevant when evaluating the quality of the decision itself. In five different experiments, Baron and Hershey demonstrated the existence of such an outcome bias (1988). In later articles, they are quick to emphasize that outcomes can affect decisions in normative ways and that their original research met stringent requirements for demonstrating the nonnormative use of outcomes (Hawkins & Hastie 1990; Hershey & Baron 1992). Such ambiguities were present, for example, in research on the similar phenomenon of *hindsight bias*, according to which individuals counted good outcomes as one of the criteria for evaluating the decisions made by other people (Zakay 1984). Outcome bias and hindsight bias are perhaps subsets of the more general phenomenon of *contaminating effects* according to which almost any known information tends to influence decision making (Chapman & Johnson 2002).

I am not aware of any research on the evolution of such contaminating effects. As Baron and Hershey note, outcomes are only irrelevant in certain peculiar situations which are probably novel to our evolved psychology. This suggests that Kahneman and Tversky's explanation of cognitive biases as computationally and metabolically efficient alternatives to more accurate decision processes may explain such contamination effects. Stanovich and West report a negative correlation between outcome bias expression and a cognitive ability composite and conclude that the bias reveals the computational limits of the human brain (1998). This finding lends further support to Kahneman and Tversky's explanation. Finally, if one considers Buss and Haselton's three criteria for EMT, then that theory does not seem to explain outcome bias. In contrast to situations involving uncertainty which are repeated throughout history (is that a snake in the grass?), the outcome bias experiments posit novel situations with perfect knowledge.

The contamination effects, including outcome bias, suggest an elegant solution to the problem of how people can forget that agents overlooking the View from Nowhere do not yet have any reason to prefer their later characters over any other. People have a natural tendency, in accordance with these biases, to take into account almost any known information—including the information of what characters they were eventually given and with which they have grown familiar. Given these contamination effects, the endowment effect and mere exposure effect become available to give people a subtle reason for preferring their later selves from any other possible character. In this way, people may jump out of the View from Nowhere and into the View from Somewhere. Certain other biases, however, provide further means by which people can escape the View from Nowhere's paralyzing sting.

One cognitive bias is so important to social psychology that researchers have labeled it the *fundamental attribution error* (FAE) (Ross 1977). According to the FAE, people overemphasize dispositional influences on observed behavior and underemphasize situational influences on this same behavior. In the classic experiment performed by Jones and Harris, observers listened to a person read an essay that either supported or opposed Cuba's president, Fidel Castro (1967). The observers were told either that (i) the person reading the essay had chosen which side to take or (ii) that the reader had been told which side of the position to take. In conformity with the systematic model of dispositional inference published by Jones and Davis (1965), observers inferred that the essay reader had the same attitudes towards Castro as expressed by the essay. But in the alternative scenario, and contrary to predictions, the observers *still* made similar inferences. In the decade following the classic experiment by Jones and Harris, the FAE

“was replicated under a variety of circumstances that ruled out some of the more obvious artifactual explanations” (Gilbert & Malone 1995).

Without making a single reference to evolution, Gilbert and Malone speculate that the FAE exists because it is “easy” and “may have few unfavorable consequences and many favorable consequences” (1995). But other researchers have noted that the FAE, unlike outcome bias, lends itself to analysis according to EMT. In particular, Haselton and Nettle (forthcoming), Haselton and Buss (2003), and Andrews (2001) have approached from an evolutionary perspective the same “consequences” that Gilbert and Malone mentioned. Andrews suggests that, when our ancestors evaluated the trustworthiness of suspicious others, the costs of an erroneous inference of guilt were relatively low and the costs of an erroneous inference of innocence were often high. This cost differential could enable a bias to evolve such that humans err on the side of assuming that behavior reflects disposition. Furthermore, as Andrews notes, correspondent causes of behavior are few but potential noncorrespondent causes of behavior are potentially infinite in number. So I would add that correspondence is *empowering* in a way that noncorrespondence is not: one can manage problems with familiar agents but one cannot manage problems with one from an unknown number of potentially infinite causes. In this way, a finding of noncorrespondence, although reasonable, may be *worthless*. So the FEA, according to EMT, may reflect how relatively empowering and worthless findings of correspondence and noncorrespondence are. In accordance with Kahneman and Tversky’s theory, the FAE may also reflect our cognitive limitations when evaluating the behavior of others. So, given our cognitive limitations, the heuristic “the person’s behavior follows from her disposition” may be

preferable to a less efficient and more comprehensive evaluation of potentially infinite number of noncorrespondent causes. Finally, in the face of so many potential noncorrespondent causes, the FAE might just reflect another cognitive bias: the need for closure, whereby individuals have a need to form and freeze a definite position on any given issue (Kruglanski & Webster 1996). Alternatively, the FAE may be a result of the ambiguity effect, whereby people prefer options for which a favorable outcome is known over options for which a favorable outcome is unknown (Ellsberg 1961).

Considering that the free will problem is about given-ness, the FAE becomes relevant to the free will problem when one observes another performing not just any action but the particular action of *forming one's character*. According to the FAE, observers would overestimate the importance of the person's character, and underemphasize the importance of situational constraints, when explaining the other's character formation. In the limiting case, one would regard the characters of others as forming themselves *ex nihilo*. The FAE allows the agent, looking onto the View from Nowhere, to exist in uneasy tension between its currently blank and future selves.

In enabling people to jump out of the View from Nowhere, the FAE may work in concert with another cognitive bias: *anthropomorphic bias*. Anthropomorphic bias is the tendency of people to attribute humanlike qualities to non-human things. Although this bias has not received enough attention (Guthrie 1993: 57), classic experiments have demonstrated its existence (Dennis 1953), and it is often discussed in the literature on religion (Guthrie 1993; Barrett & Keil 1996; Caporael 1986). In the religious context, Guthrie notes that Bacon (1960: 51-52), Hume (1957: 29), Darwin (1871: 67), Nietzsche (2001: 78), and Freud (1989) explained religion, in part, as anthropomorphism. The

tendency is so common that it has its own name in literary criticism: the pathetic fallacy (Abrams 1993). As Hume observed:

“There is an universal tendency among mankind to conceive all beings like themselves, and to transfer to every object, those qualities, with which they are familiarly acquainted, and of which they are intimately conscious. We find human faces in the moon, armies in the clouds; and by a natural propensity, if not corrected by experience and reflection, ascribe malice or good-will to every thing, that hurts or pleases us. Hence the frequency and beauty of the prosopopoeia in poetry; where trees, mountains and streams are personified, and the inanimate parts of nature acquire sentiment and passion. And though these poetical figures and expressions gain not on the belief, they may serve, at least, to prove a certain tendency in the imagination, without which they could neither be beautiful nor natural.”

Guthrie notes that many of these thinkers cited the importance of familiarity and comfort in explaining religion as anthropomorphism. According to these theories, people distrust the unknown (this observation anticipates the discovery of the mere exposure effect), and so they prefer to believe God is like them. Although Guthrie claims that such theories have “some small truth,” he prefers his own “cognitive, evolutionary, and game-theoretical” one. According to Guthrie, guessing that something is human is a “good bet.” In explaining his theory, Guthrie suggests two elements of EMT: the uncertainty of the situation (“in a complex and ambiguous world our knowledge always is uncertain”) and the relative costs of errors (“if we are right, we gain much, while if we are wrong, we usually lose a little”) (1996). His theory would also satisfy the third element: distinguishing between correspondent and noncorrespondent causes of behavior would have had “recurrent effects on fitness over evolutionary history.” Indeed, Haselton and Nettle observe that “Guthrie uses error management logic” to explain anthropomorphic bias. Both Guthrie (2002), as well as Atran and Norenzayan (2004), have later elaborated on the use of EMT to explain anthropomorphism.

Anthropomorphic bias becomes relevant to the free will problem when one considers this long tradition of homunculus theories. In particular, homunculus theories are an expression of the anthropomorphic bias with respect, not to other objects such as clouds or animals, but to *subdivisions* of the human mind. Recall the agent, in uneasy tension between its currently blank and future selves, overlooking the View from Nowhere. A rational observer would regard this agent's character as blank. Tension might arise, however, if an observer projects anthropomorphic features onto this agent. In particular, an observer might project anthropomorphic goals and values. So, from amongst the vast space of characters (the overwhelming majority of which will be non-human), the agent might prefer a human character. This homunculus theory enables and strengthens the FAE in the context of one's character formation by positing a preexisting character which can form itself *ex nihilo*.

There is strong evidence that just such an anthropomorphic bias has been motivating some of the parties in the free will debate. Kane notes that “[s]ome philosophers reified the Will as a mysterious *homunculus* within the agent” (2001); Claxton writes that “[cognitive scientists] have been at pains to do away with the *homunculus* which is so often concealed within the folk theory of personhood of which free will is such an integral part” (2000: 104); and Walter observes that “[o]ur everyday explanations rest on the assumption that the essence of an agent is something at its core, conceived as a soul or *homunculus*” (2001). Dennett, a compatibilist, criticizes Kane, a libertarian, for committing this homunculus fallacy (2003: 123); similarly, I have criticized Fischer, a prominent compatibilist, for also thinking of human agency in this way (forthcoming).

Another cognitive bias which may aggravate the FAE is the asymmetrical attribution of blame. Research shows that observers of immoral actions are more likely to infer correspondence between the actions and the actor's character (Reeder & Spores 1983). In contrast, observers of moral actions are more sensitive to the situational constraints which may have caused this behavior. These results are consistent with other research showing that people identify words implying social costs than words implying poor skill or positive qualities (Ybarra et al. 2001). The result is that the FAE may be especially insidious in the context of observing immoral, rather than moral, behavior.

Asymmetrical attributions of blame may reach their zenith in the process of *demonization* (Ellard et al. 2002). Demonization is an unwillingness to empathize with another such that the person regards this other as *evil*. Despite correcting for any potential differences in personality, by having the same person write both a story as a perpetrator and a story as a victim, Baumeister, Stillwell, and Votman found significant differences between the two kinds of stories: perpetrators recalled much shorter time spans, victims regarded the actions of perpetrators as more inexplicable, and victims regarded the perpetrator's actions as more harmful (1990). One might worry that these findings are only relevant to attributions of evil but Baumeister and Vohs are quick to suggest that demonization admits of degrees and so may be a common feature of our responsibility practices:

“To be sure, our research sample consisted of many everyday conflicts and misdeeds, few of which were sufficiently important to qualify for the grandiose term *evil*. Our assumption, however, is that similar processes operate in everyday transgressions as in large-scale misdeeds, and that if anything, the gap between victim and perpetrator would probably be even larger in horrendously evil events than in petty, everyday conflicts.” (2005: 87)

In the context of Nagel's *The View from Nowhere*, the second of these differences is the most relevant. The researchers reported that although "[b]oth victims and perpetrators distorted their stories—and to almost identical degrees," "the weight of the evidence tends to be closer to the perpetrators' accounts" (2005: 89-90). This is so because "[p]eople rarely attack for no reason" even though the "perpetrator's motives are often opaque to the victim" and "victims cannot or will not see this perspective" (2005: 88-89). These results suggest how the process of demonization puts the other into the View from Nowhere—where mitigating circumstances disappear.

In explaining the cause of asymmetrical attributions of blame and demonization, researchers have noted that these biases lend themselves to EMT analysis (Haselton & Nettle, forthcoming). Such analysis requires that the respective costs of false positives and false negatives differed over evolutionary history. But asymmetric attributions of blame reflect, by distinguishing between moral and immoral actions, these differing costs between false positives and false negatives. For example, the cost of mistakenly assuming that a murderer is a friend may have been more costly than mistakenly assuming that a friend is a murderer.

The FAE, together with anthropomorphic bias and asymmetrical attributions of blame, provides a more plausible context in which the endowment effect and outcome bias give people an irrational reason to have preferred their later characters from amongst the space of possible characters. These four cognitive biases form a potent mixture which inclines one to regard others as possessing more control over their lives than actual control; we regard others as possessing control over their lives which approaches

novelist* control. Together, these biases inoculates people against the View from Nowhere’s paralyzing sting—allowing them to jump into the View from Somewhere.

This section has focused upon how people attribute responsibility to *others*. In Nagel’s famous essay, he calls this the problem of *responsibility*. But Nagel noted another problem—which is perhaps more disturbing—involving how we attribute freedom to *ourselves*. Nagel called this the problem of *autonomy*. Although the two problems sometimes intersect, distinct cognitive biases may account for inflated senses of responsibility and autonomy. The following section (III) will focus upon this second problem of autonomy. In particular, it will explore the “positive illusions” which inflate one’s sense of control and suggest a novel explanation for how these illusions evolved.

III. FALSE ADVERTISING AND THE POSITIVE ILLUSIONS

Humans are vulnerable to a variety of positive illusions about *themselves*. In a seminal article, Taylor and Brown reviewed and integrated the research on such positive illusions and concluded that they be a sign, not of mental illness, but of mental health (1988). Taylor and Brown later defended (1994) their work against, and their conclusions seem to survive, a critique by Colvin and Block (1994).

Taylor and Brown divide the positive illusions into “unrealistically positive self-evaluations, exaggerated perceptions of control or mastery, and unrealistic optimism.” First, consider the tendency of humans to evaluate themselves in irrationally positive ways (Greenwald 1980). This tendency is related to the Lake Wobegon effect, whereby the average person rates herself as above-average on various measures—which is logically impossible. As Taylor and Brown note, this positive illusion reveals itself:

“over a wide range of traits” (Brown 1986) and abilities (Larwood & Whittaker 1977), in asymmetrical recall of valenced information about the self (Silverman 1964), in a tendency to view one’s own group as better than other groups (Tajfel & Turner 1986), and in comparisons between self-evaluations and evaluations by others (Lewinsohn et al. 1980).

These positively biased self-assessments may complement the endowment effect in explaining the illusion of free will. The endowment effect would allow one, given no consensus on values, to value what one has been given. In contrast, positively biased self-assessments would allow one, given a consensus on values, to feel that one has maximized the traits that would further those values. But it is not clear that the two biases are so distinct or, if they are, that the former has no role to play in the free will problem. Consider the measure, not of traits such as warmth or integrity whose value is uncontroversial, but of having a good character in general. Unlike being more or less warm, or having more or less integrity, there seems to be little reason to prefer being the kind of person one is to being someone else. For example, there is little reason to prefer, within a certain healthy range, being more or less sensual or more or less introverted—to regretting the accidents of birth which have made us one kind of person rather than another. Positively biased self-assessments may, however, in concert with the endowment effect, incline people to feel that they have better characters, even in this controversial sense, than others—allowing people to jump out of the View from Nowhere into the View from Somewhere.

The relevance of another positive illusion to the free will problem is more obvious—the illusion of control. Taylor and Brown trace the literature on this illusion to

Langer's classic studies in the context of gambling (Langer 1975; Langer & Roth 1975). In those experiments, Taylor and Brown showed how people feel they have more control over situations than actually have. For example, people with a light switch overestimate the amount of control they have over whether a semi-randomly flashing light turns flashes. Similarly, a wide body of literature demonstrates the same positive illusion in situations that are heavily determined by chance (Crocker 1982).

Taylor and Brown do not mention another cognitive bias which may also belong to the positive illusions—the trait ascription bias. According to the trait ascription bias, people regards themselves as variable in terms of personality, behavior and mood while regarding others as being less varied and more predictable. In the original study on this bias, Kammer had 56 undergraduates rate the behavioral consistency of themselves and their friends (1982). Kammer discovered the undergraduates rated their friends' behavior as more consistent than their own. To the extent that predictability is undesirable, the trait ascription bias may represent another positive illusion.

Taylor and Brown speculate that the positive illusions are adaptive because they promote mental health by: increasing happiness or contentment, enabling social bonding, and increasing the capacity for creative or productive work. In particular, they argue that the positive illusions can stimulate creative or productive work by facilitating intellectual functioning and increasing motivation or persistence. In support of this claim, Taylor and Brown note that “[p]ositive conceptions of the self are associated with working harder and longer on tasks (Felson, 1984); perseverance, in turn, produces more effective performance and a greater likelihood of goal attainment (Bandura, 1977; Baumeister, Hamilton, & Tice, 1985; see also Feather, 1966, 1968, 1969)” (1988). Although Taylor

and Brown discuss how the positive illusions may be adaptive they do not speculate as to how such illusions evolved.

Haselton and Nettle note that Taylor and Brown's theory, like Guthrie's theory of religion as anthropomorphism, "tacitly contains an error management argument" (Haselton & Nettle, forthcoming). In particular, Haselton and Nettle seize upon Taylor and Brown's emphasis upon motivation and persistence. Situations in which individuals must decide whether to stop or persist in their efforts may satisfy the three elements of EMT: they involve uncertainty (if I continue, will I eventually succeed?), the decisions had repeated effects upon fitness throughout history (if I succeed in wooing a mate, or killing a rival, my genes will propagate at the expense of those who fail), and the relative costs of errors or correct inferences were asymmetrical (irrational persistence often costs less than irrational quitting).

I would suggest an alternative—and perhaps more disturbing—explanation for the prevalence of the positive illusions. In my estimation, Haselton and Nettle further the tendency of psychologists after Darwin to neglect *sexual*, as opposed to natural, selection. I do not deny that EMT may explain, in part, any of the positive illusions. But an alternative explanation in the context of sexual selection seems more natural, elegant, and robust. Instead of supposing that the subtle cost-benefit analysis of certain decisions would satisfy the requirements for EMT, one can simply note how irresistible the temptation will be for organisms, when advertising to mates, to *lie*.

It is uncontroversial that people often lie in the mating context. For example, "people report being deceived by friends about mating rivalry more often than they themselves report engaging in deceit about rivalry, and women more than men deceive

each other about how sexually experienced and promiscuous they are” (Bleske & Shackelford 2001). Similarly, Buss cites research (Keenan et al. 1997) showing that “men motivated to seek casual sex frequently attempt to deceive women about their commitment, social status, and even fondness for children—domains of deception about which women are well aware” (Buss 2003: 283). These results are consistent with widespread deception in the plant and animal kingdoms. Searcy and Nowicki note that, although some have expressed doubt about deception in the animal kingdom, these doubts have been invalidated by adopting a functional definition of deception and by focusing upon individual selection instead of group selection (2005: 219). Buss notes two vivid examples: some male insects give gifts to female insects but take them back after mating (Thornhill & Alcock 1983) and some orchids mimic female wasps such that male wasps attempt to copulate with them; the male wasps fail to compliment but succeed in carrying pollen (Trivers 1985). These results are also consistent with the principle, from communications theory, that signalers have been under selection pressure to manipulate, even through deception, receivers (Krebs & Dawkins 1984). In this context, Trivers observes that deception is “a parasitism of the preexisting system for communicating correct information.” (Buss 2003: 105).

One difference between the positive illusions and deception in other contexts immediately presents itself: the usual form of deception involves knowledge of one’s deceit and the intent to deceive whereas humans experiencing positive illusions *believe their lie*. So the positive illusions involve not just deception but *self-deception*. One must therefore explore the evolutionary origin of self-deception in order to explain the evolutionary origin of the positive illusions.

Trivers first suggested how self-deception might evolve in his classic forward to Richard Dawkin's *The Selfish Gene*:

“The arguments themselves extend in many directions. For example, if (as Dawkins argues) deceit is fundamental in animal communication, then there must be strong selection to spot deception and this ought, in turn, to select for a degree of self-deception, rendering some facts and motivates unconscious so as not to betray—by the subtle signs of self-knowledge—the deception being practiced. Thus, the conventional view that natural selection favors nervous systems which produce ever more accurate images of the world must be a very naïve view of mental evolution” (1976).

Trivers' fundamental insight is that the best liars believe their own lies, because they cannot betray their knowledge of the deceit, and so there would be selection pressures upon deceivers to deceive *themselves*. He later explored the evolutionary origins of self-deception in more detail (2000). First, Trivers sympathizes with the Taylor and Brown's hypothesis that the positive illusions bring intrinsic benefit to the individual. Second, Trivers argues that self-deception can express itself in five ways: denial of ongoing deception, unconscious modules involving deception, self-deception as self-promotion, the construction of biased social theory, and fictitious narratives of intention. There seems to be some overlap between these kinds of situations; the first three may combine to form, amongst other things, the positive illusions. People may believe that are better than they actually are, or have more control over their lives than they actually do, to better convince others of these falsehoods.

Indeed, Trivers anticipates just this explanation for the positive illusions. He observes that one source of self-deception “has to do with self-promotion, self-exaggeration on the positive side, denial on the negative, all in the name of producing an image that we are ‘benefective,’ to use Anthony Greenwald's apt term, toward others”

(2000: 117). But one does not need the idea of being “beneffective,” and inviting reciprocal altruism, in order to explain the positive illusions. Trivers does not mention a more fundamental form of cooperation amongst the animal and plant kingdoms, including homo sapiens—sex.

Being above average on any given positive trait is *sexy*. This is especially true considering that people may adapt their preferences to the conditions of the mating marketplace. In the class study on the *contrast effect*, Kenrick and Gutierrez found that men rated an average woman as less attractive after watching *Charlie’s Angels* (1980). Buss discusses several findings that followed this classic experiment and shows how they are consistent with an evolutionary theory of happiness (2000). For example, people disregard objective measures of well being humans are on a “hedonic treadmill” (Diener et al. 1999) and because “*differential* reproductive success is the engine of the evolutionary process” (Buss 2000). Buss describes how an evolutionary theory of happiness reveals that we are on a treadmill such that supposed increases in well-being, like winning the lottery, or the increase in standard of living throughout the twentieth century, do not cause long term increases in happiness. Being a rich supermodel is less attractive if everyone is a rich supermodel. In the mating marketplace we are all graded on a curve.

Similarly, control over one’s life is *sexy*. From the perspective of *female choice*, one’s good intentions are worthless to a woman if one is powerless to effect them. Buss details how women find economic capacity, social status, ambition and industriousness, dependability and stability, intelligence, size and strength, and good health to be

attractive (Buss 2003)—all indicators of control over one’s life. As Henry Kissinger observed, “power is the ultimate aphrodisiac.”

There is evidence that such power is important to libertarian conceptions of free will. For example, Kane defends a theory of event-causal libertarianism according to which free will involves “ultimate control” (1999) or “ultimate dominion” (1985) over one’s life. Other theories of agent-causal libertarianism characterize humans as godlike unmoved movers. (O’Connor 2005). If power is the ultimate aphrodisiac, then libertarian free will would render one extremely sexy.

The case for the attractiveness of creative variability is more subtle than the case for the attractiveness of control over one’s life. This case is largely based upon Miller’s treatment of the subject (2001). Miller begins by observing that the behavior of simple organisms is often predictable; he further notes that, to the extent that such organisms are prey, this predictability is problematic. Miller describes how Von Neumann discovered that the best strategy in certain games from game theory involve randomness. Later, Fisher and others discovered examples of just such randomized strategies in the animal kingdom. These considerations lead Miller to conclude that this overall strategy, which Driver and Humphries called protean behavior, is a fundamental adaptation found in almost all mobile prey. Miller speculates that protean behavior expresses itself in humans through *creativity*. Such creativity is strongly correlated with reproductive success (Nettle & Keenoo 2005).

These last observations become relevant to the free will debate when we consider just how important creativity and unpredictability are in that context. One traditional definition of free will is the ability or power “to do otherwise”; a large literature explores

just what this power entails and whether moral responsibility requires it. Moreover, one virtue of libertarianism is that it renders human decisions *unpredictable* by definition. Amongst compatibilists, Fischer has characterized the value of free action as the value of *creative self-expression* (2005; 2006a).

It is important to note that, although the concept of free will implies both control and unpredictability, perhaps expressed in the illusion of control and trait ascription bias respectively, these may represent two entirely different—and mutually exclusive—features of free will with distinct evolutionary origins. For example, Richard Double claims that no libertarian account of free will can jointly satisfy the following three requirements: control, rationality, and the ability to choose otherwise (1991, chapter 8). Putting aside the worry about rationality for a moment, one can note the tension between the desire for control and the desire for unpredictability (which would correlate with the ability to choose otherwise). Each seems to do violence to the other: greater control over one's life ensures that one's behavior will follow from one's character—enhancing predictability—and greater unpredictability ensures that one's behavior does not follow from one's character—undermining control. This tension may reflect the distinct evolutionary origins of these desires which have been coupled together within the concept of free will.

Those who study the free will problem must remember that throughout history there have been tremendous selection pressures upon humans to become better than average, to obtain control over their lives, and to express creative variability. Furthermore, to the extent that not everyone has been able to attain these lofty ideals, there have been tremendous selection pressures for humans to *fake* them. In the same

way that a man might lie about being married in a singles bar, so too might a man lie about having more control over his life than he actually does. One can consider this phenomenon using Darwin's notion of *female choice*. Throughout the ages, males who succeeded in convincing females that they had more control over their lives than they actually had (perhaps approaching novelist* control) may have enjoyed more reproductive success than males who admitted to just having actual control over their lives. Similarly, males who convinced females that they were more creative than they actually were may have enjoyed more reproductive success. Furthermore, Trivers shows how one of the best ways for males to succeed in their deception was to *believe their lies*.

By analogy, suppose that a corporation such as Nike engages in behavior that is unattractive to the consumer. For example, suppose that some accusations are correct and Nike employs people in sweat shop conditions for unreasonable pay. Consider Nike employees responsible for marketing the company's brand. Of course, the marketing employees would not advertise the sweatshop conditions. Indeed, they might even lie about them. It is not unreasonable to suppose that such employees would be so committed to their jobs that start to believe their own lies—and this makes them better liars. Finally, suppose that a group of marketing employee meets a person in a wheelchair. These marketing employees might tell the handicapped person that Nike does not employ workers in sweatshop condition even though this person is unlikely to ever buy athletic wear. The employees might do this because they are protecting a brand name. Handicapped persons are not islands; they may spread word about Nike to potential customers even if they themselves will never purchase Nike products. Marketing employees are promoting a brand and cannot afford to make subtle and

dangerous distinctions between who may and may not learn of information damaging to that brand. By analogy, the positive illusions may have evolved because they allow people to promote themselves, not only to potential mates, but to society in general—and the best liars believe their own lies.

These first two sections (**II** and **III**) have explored how certain cognitive biases might have evolved because they were *adaptive*. These adaptive biases would give one direct reasons to regard others (**I**) and the self (**II**) as having more control over their lives than is actual—having control over their lives that approaches novelist* control. Such biases might have been adaptive in accordance with one or more of the following: a heuristics and biases theory, error management theory, or false-advertising through self-deception. The next section (**IV**) will argue that evidence of one last group of cognitive biases would only provide indirect reasons for regarding others and the self as possessing more control over their lives than is actual; nevertheless, these biases lend strong support to a great philosopher’s error theory of belief in free will. The section will further argue that such biases evolved as adaptations or in other non-adaptive ways.

IV. IRRATIONAL OPTIMISM

The three categories of positive illusions discussed by Taylor and Brown involve biased evaluations of the self. The previous section considered two of these three categories: irrational self-esteem and the illusion of control. This section will focus upon the third positive illusion: irrational optimism. Irrational optimism is the central focus on this section but it is itself just a subset of the remaining cognitive biases which favor belief in free will. The compliment of the positive outcome biases within this larger set is

composed of biases which cause *belief inertia* given the popularity of belief in free will: the bandwagon effect and status quo bias. From this perspective, the popularity of free will presents another obstacle to rational belief in that power.

According to the bandwagon effect, people regard popular ideas as more accurate than unpopular one. The effect is discussed in a classic article (Lee & Lee 1939), in the voting context (Allport 1940; Simon 1954: 245-253; Carter 1959), and has been recently repeated in that context (Mehrabian 1998; Marsh & O'Brien 1989). For example, Mehrabian reports two experiments which showed how bogus polls significantly affected voting, "supporting the bandwagon effect."

One implication of the bandwagon effect is the "spiral of silence." According to Noelle-Neumann, individuals may value belonging with the majority more than they value holding their own opinion (1974). The result is a "spiral of silence" which magnifies the prominence of majority views and suppresses minority ones. A recent meta-analysis concluded that the "spiral of silence" effect is small but significant (Glynn et al. 1997).

Related to the bandwagon effect is another cognitive bias which would favor belief in free will: status quo bias. Status quo bias is the irrational preference people express for things staying the same (Samuelson & Zeckhauser 1988). For example, Kahneman, Knetsch, and Thaler cite the following example of status quo bias (1991). Consumers of a California power grid were divided into two groups such that one had more reliable service than the other. The consumers were then surveyed about their preferences. Both groups selected their status quo as the most desirable option. Kahneman, Knetsch, and Thaler report that subjects in another experiment expressed

similar preferences for the status quo when considering whether to adopt two kinds of automobile insurance (Johnson et al. 1993).

I know of no literature on the evolution of the bandwagon effect and status quo bias. But one might speculate as to how these biases evolved. First, one might subject them to EMT analysis. The question of whether to join the majority or not would be a recurrent event throughout history that affects an individual's fitness. Furthermore, the relative costs of false positives and false negatives, or correct positives and correct negatives, might vary. For example, the cost of mistakenly introducing an unpopular idea to the group, and incurring its wrath or disapproval, might outweigh the cost of mistakenly suppressing the idea and preserving the status quo. This is especially true considering that the cost associated with announcing new ideas depends, not just on how accurate the new ideas are, but how much people have invested in old ideas. Nobody likes to admit they were wrong and people will inflict fitness costs on others who threaten to expose their errors. The analysis here would be similar to that of the mere exposure effect, according to which selective pressures might favor a conservative attitude in an uncertain world. This variance would put selective pressure upon the species to evolve a cognitive bias tuned to avoid the greater costs—at the expense of accuracy. In considering the evolution of these biases, one might also note that Kahneman, Knetch, and Thaler regard status quo bias as an implication of the endowment effect and loss aversion (whereby people express undue dislike for losing their possessions). To the extent that their theory is correct, the analysis of how those other biases evolved would apply to status quo bias as well. Furthermore, the bandwagon effect and status quo bias

bear a resemblance to the contrast effect and to that extent the discussion of how the contrast effect evolved would also apply here.

These belief inertia biases only show that people find novel ideas *unpalatable* but not that they find them *unlikely*. To complete this section's case for the remaining cognitive biases favoring belief in free will, one must return to the *positive illusions* which relate, not to self-evaluations or the illusion of control, but to the likelihood of good things happening. These biases include *belief bias*, the *confirmation bias*, and the *choice-supportive bias*. Each of these belongs to the set of positive outcome biases because they incline people to regard certain positive outcomes as more likely than negative ones.

First, consider the third positive illusion: the *valence effect*. According to the valence effect, people overestimate the likelihood of good things happening to them and underestimate the likelihood of bad things happening to them. This effect seems to be quite general. For example, Rosenhan and Messick reported that, all else being equal, subjects considered turning over a card with a smiling face on it to be much more likely than turning over a card with an angry face on it (1966). They found no effect during an experiment that used the neutral images of big and little kangaroos. Similarly, in their review of the positive illusions, Taylor and Brown cite the following studies in support of the valence effect:

“People estimate the likelihood that they will experience a wide variety of pleasant events, such as liking their first job, getting a good salary, or having a gifted child, as higher than those of their peers (Weinstein, 1980). Conversely, when asked their chances of experiencing a wide variety of negative events, including having an automobile accident (Robertson, 1977), being a crime victim (Perloff & Fetzer, 1986), having trouble finding a job (Weinstein, 1980), or becoming ill (Perloff & Fetzer, 1986) or depressed (Kuiper, MacDonald, & Derry, 1983), most people believe

that they are less likely than their peers to experience such negative events.” (1988)

The valence effect may account, in subtle ways, for a variety of cognitive biases related to belief inertia. First, *belief bias* is the phenomenon whereby subjects reject valid arguments with unbelievable conclusions and accept invalid arguments with believable conclusions (Evans et al. 1983). Belief bias is an error of deductive reasoning and may be an example of the larger phenomenon of *belief perseverance*. Belief perseverance is the cognitive bias which expresses itself as the irrational maintaining of beliefs in the presence of contrary evidence. For example, Anderson, Lepper, and Ross found that after subjects formed beliefs they would discredit contrary evidence and their beliefs would remain resistant to change (1980).

A subset of the valence affect may be *confirmation bias*. Confirmation bias enables belief perseverance by inclining people to seek confirmation of their preconceptions at the expense of alternative hypotheses. In the classic experiment, Wason presented subjects with three numbers (1960). The subjects were asked to determine the appropriate rule governing the sequence by generating their own sequences and submitting them to the researcher for feedback. Wason observed that subjects tended to submit only positive examples which would confirm their suggested rule and did not submit negative examples which would disprove it. He called this tendency the confirmation bias. A recent review of research on the confirmation bias concluded that “[i]n the aggregate, the evidence seems to me fairly compelling that people do not naturally adopt a falsifying strategy of hypothesis testing” (Nickerson 1998).

A third positive illusion related to belief is *choice-supportive bias*. Choice-supportive bias expresses itself in the tendency of humans to remember their chosen

options as having more positive attributes than they do in fact. For example, Mather, Shafir, and Johnson gave subjects a choice between two options and later had these subjects evaluate their choices (2000). The subjects reported choice-supportive bias by tending to attribute, both correctly and incorrectly, more positive attributes to their chosen options than to their alternatives.

Earlier, this article suggested that the belief inertia biases may evolved through EMT in a way analogous to the evolution of one or more of the mere exposure effect, the endowment effect, loss aversion, and the contrast effect. In contrast, the positive illusions with respect to belief may have evolved in a way analogous to the other positive illusions. As discussed in the previous section, the positive illusions may have evolved because they are *sexy*. Similarly, the valence effect, belief bias, confirmation bias, and choice-supportive bias may represent false-advertising which benefits the individual at the expense of others. Furthermore, as Trivers argued, one of the best ways to lie is to believe one's own lie.

These cognitive biases about both the palatability and likelihood of altering beliefs in the face of contrary evidence favor belief in free will in three distinct ways. First, to the extent that people regard free will as *desirable*, the valence effect would incline them to inflate the likelihood that free will exists. There is no denying that people tend to regard free will as a good thing. For example, considering the aspect of free will involving control, it is important to must note that feeling in control is an essential ingredient to people's happiness (Larson 1989). With respect to the aspect of free will involving unpredictability, one must note how disturbing is the idea that another might predict or design one's entire life. Given the importance of control and unpredictability

to feelings of well-being, one would expect, in the context of the free will problem, the valence effect to be at its zenith. Just as people regard the possibility of turning over a card with a smiling face on it as unduly likely, so too may they regard the possibility that free will exists as unduly likely.

Second, these biases favor such belief in a *subtle* way by implying that the control mechanisms humans have over accepting and rejecting beliefs are sufficient. If humans systematically feel satisfied with the beliefs they have formed, they will feel no need to demand greater control over the beliefs they accept or reject. They may grow complacent with the control mechanisms one has over acquiring such beliefs.

In response to the suggestion that these biases are related to the free will problem in this subtle way one might insist that the free will problem is associated with control over *decisions* and not *beliefs* (Strawson 1986). This is true. One demands freedom over whether to choose to turn left or right and not whether to believe that the sun is shining. Belief in this latter fact is in one sense *coerced* upon humans. Yet they do not *resent* this coercion and the demand for control over whether the sun shines may be unreasonable. But this distinction is not as clean as its defenders might suppose. One can distinguish between decisions and beliefs but one must remember that beliefs affect decisions—so a problem with the former can *infect* the latter. This is true in the case of our beliefs about the physical world. One can choose to go left or right but this decision is constrained by countless beliefs about the world. One cannot go left if there is a wall opposing you. The critic might reply that even if one cannot go left one can decide to go left—the will can remain free while the body is enslaved. But the question immediately presents itself: *why* would one choose to go left? This raises the more obvious and disturbing worry: beliefs

about *value*, and not about the physical world, can infect decisions too. One maintains beliefs about value just as much as one maintains beliefs about the world but these values place even more intimate and tight constraints upon our decisions. A strong willed individual cannot decide to go left if she does not feel that going left furthers her values—so freedom over decisions may be worthless without freedom over beliefs. In this subtle way, these cognitive biases may render humans complacent with the control mechanisms they have over their belief acquisition.

But these cognitive biases may favor belief in free will in another obvious and disturbing way: that free will exists is *itself* a popular belief and so these biases will not only lead to complacency over the control mechanism they have over belief acquisition; they will also lead to complacency about the particular belief in free will.^{vii} This difficulty is just a result of the doctrine's popularity and adds to the difficulty posed by other cognitive biases.

Of course, a consensus exists about many ideas and most of them do not need revision. The fact that an idea is popular does not, by itself, suggest that it is false. But it does suggest that, if the idea is false, then realizing this fact and discarding the old belief will be difficult. Not all popular ideas are right. Geocentrism, creationism, and racial inferiority only died—to the extent they have died—in the face of extraordinary opposition. Galileo, Darwin, and Martin Luther King, amongst others, suffered or died for being prophets. They were victims of the bandwagon effect, status quo bias, the valence effect, belief bias, confirmation bias, and choice-supportive bias. It is important to remember that fact when considering the possibility of other popular falsehoods and potential prophets—the existence of free will has been questioned by Nietzsche (1886:

§21), Spinoza (1985: proposition 48), Darwin (1987: 608), Einstein (1994: 262) and Russell (1927).

When one considers the philosophical literature one detects subtle evidence that irrational optimism is motivating at least some of the views in this area. For example, Dennett's defends compatibilism by focusing just upon the "varieties of free will worth wanting"—at the expense of those varieties which are not worth wanting. In defending non-realism about free will, Pereboom draws attention to Dennett's rhetorical strategy:

"I think it is important to distinguish whether we are free in the sense required for moral responsibility from whether it is valuable to be free in this sense. Here I am resisting a trend initiated (to my knowledge) by Daniel Dennett. His attempt to recast the debate in terms of the question, 'What is free will such that we should want it?' potentially confuses two issues: Do we have the sort of free will required for moral responsibility? And do we want the sort of free will required for moral responsibility? It could be, for instance, that we are free in the sense required for moral responsibility, but since being free in this sense is not especially valuable to us, we would not want it much. It is important to frame the issue so as to make conceptual room for views of this type." (2001: xxii)

Pereboom suggests that Dennett is confusing, and perhaps conflating, the question of whether free will exists and the question of whether it is valuable, such that free will's value renders its existence more likely. This would be irrational optimism. More generally, Tamler Sommers has observed that "[p]hilosophers who reject God, Cartesian dualism, souls, noumenal selves, and even objective morality, cannot bring themselves to do the same for the concepts of free will and moral responsibility" (forthcoming).

Nietzsche suggests how one last subset of irrational optimism explains the widespread belief in free will. In response to this suggestion, one might feel tempted to dismiss Nietzsche as more of a proto-postmodern author of literature than as an analytic philosopher or cognitive psychologist. In that case it is important to remember that,

according to Leiter and Knobe, support for Nietzschean moral psychology remains robust in the post-Freudian world (forthcoming). Leiter and Knobe show how modern psychology better supports Nietzsche's views on human moral capacity than those of Aristotle or Kant. In particular, they argue that the more important factor to the exercise of one's moral capacity is not childhood upbringing (contra Aristotle), nor conscious obedience to moral principles (contra Kant), but heritable psychological traits.

Although Leiter and Knobe consider how modern research supports Nietzsche's views on these aspects of moral psychology, they do not address Nietzsche's error theory of belief in free will. Elsewhere, Leiter expands upon Nietzsche's theory of how the masses fit embrace the ascetic ideal despite striving to maximize feelings of power. This expansion, which one finds in Nietzsche's *On the Genealogy of Morals*, provides a *context* for Nietzsche's error theory of belief in free will, which one finds in his *Twilight of the Idols*. Leiter (2004) sums up Nietzsche's larger theory of how the masses embrace the ascetic ideal as follows:

- I. Suffering is a central fact of the human condition
- II. Meaningless suffering is unbearable and leads to "suicidal nihilism" (GMIII:28).
- III. The ascetic ideal gives meaning to suffering, thereby seducing the majority of humans back to life, i.e., it maximizes their feeling of power within the constraints of their existential situation.

Furthermore, as Leiter notes, Nietzsche explains throughout *On the Genealogy of Morals* how the *priest* serves as a *doctor* to these masses. The priest does this in two innocent ways and one guilty way. The two innocent ways are (i) dulling the pain of this world (3:17) and (ii) engaging the masses in hard work (section 3:18). The guilty way involves stirring within the masses feelings of sin and guilt (section 3:20).

Having considered the role of priests in this Nietzschean context, one can better understand his error theory of belief in free will. Nietzsche writes that:

“Today we no longer have any tolerance for the idea of ‘free will’: we see it only too clearly for what it really is—the foulest of all theological fictions, intended to make mankind ‘responsible’ in a religious sense—that is, dependent upon priests. Here I simply analyze the psychological assumptions behind any attempt at “making responsible.”

Whenever responsibility is assigned, it is usually so that judgment and punishment may follow. Becoming has been deprived of its innocence when any acting-the-way-you-did is traced back to will, to motives, to responsible choices: the doctrine of the will has been invented essentially to justify punishment through the pretext of assigning guilt. All primitive psychology, the psychology of will, arises from the fact that its interpreters, the priests at the head of ancient communities, wanted to create for themselves the right to punish--or wanted to create this right for their God. Men were considered “free” only so that they might be considered guilty--could be judged and punished: consequently, every act had to be considered as willed, and the origin of every act had to be considered as lying within the consciousness (and thus the most fundamental psychological deception was made the principle of psychology itself).” (1968: 499)

Here Nietzsche presents the priest in a less sympathetic light. As Nietzsche’s larger explanation of how the masses embrace the ascetic ideal shows, however, the priest seeks the right to punish in order to give *meaning* to the lives of his patients. By rendering meaningless suffering into meaningful suffering, the priest maximizes the masses’ will to power. They maximize this will to power by deluding themselves that the world is *just*. But, unlike Knobe and Leiter’s work on Nietzsche’s moral psychology, Leiter’s expansion on Nietzsche’s larger theory does not consider whether modern psychology supports Nietzsche’s theory of free will as enabling belief in a just world.

There is overwhelming evidence for just such a “just world phenomenon.” Lerner and Simmons first documented this effect in a classic paper (1966) and Lerner summarizes early research in *The Belief in a Just World: A Fundamental Delusion*

(1980). As an example, consider the classic experiment by Lerner and Simmons, in which a woman is subjected to pretend electric shocks while working on a difficult memory problem. People who observed the experiment “rejected and devalued her when they believed they would continue to see her suffer in a 2nd session, and when they were powerless to alter the victim’s fate.” Similarly, Lerner reports another experiment in which “observers persuaded themselves that a fortuitously rewarded worker had performed better than his partner who was deprived, also by chance” (1965).

More recently, researchers have shown that people will most associate prior behavior with a subsequent and unrelated event when those events fit deservingness expectations (good person won the lottery; bad person injured in car accident) (Callan et al., forthcoming). Furthermore, people make the same associations when the prior event involves greater suffering than when it involves lesser suffering (person with HIV suffered more or less). Other recent articles have reviewed the abundant research which replicates this finding and suggested ways for the theory to expand (Furnham 2003; Hafer & Bègue 2005).

Although Lerner’s discovery of the just world phenomenon has produced a robust literature, it exists in tension with another line of research showing that justice is just a “tool of people’s self-interest” (2003). Lerner explains this distinction by invoking a *dual-process* theory of the justice motive (Chaiken & Trope 1999). According to Lerner, participants in abstract and emotionally neutral situations, with slow response times, feel more motivated to please the experimenters and justify their actions in terms of self-interest. But participants in concrete and affect-laden situations, with short response

times, do not have the opportunity to rationalize their decisions and so “the desire for justice may become the dominant motive guiding their thoughts and behavior.”

In support of Lerner’s dual-process theory with respect to the free will problem, a landmark study shows that the distinction he draws is particularly important to intuitions about compatibilism or incompatibilism. Nichols and Knobe gave subjects descriptions of two different situations such that one was cold and abstract and the other was emotional and concrete (forthcoming). They discovered that subject who given the cold and abstract situation reported incompatibilist intuitions but these intuitions disappeared when they were given the emotional and concrete situation. Nichols and Knobe suggest that these results may reflect either affective competence or performance errors. As others have noted, however, the affective competence model seems too charitable to compatibilists. Instead, Nichols and Knobe’s performance error model presents the intriguing possibility that affect distorts rationality and motivates compatibilism—creating a “moral illusion.”

Lerner’s discovery of the just world phenomenon provides evidence of just the sort of cognitive dissonance Nietzsche described almost a century earlier. If people alter their beliefs in order to preserve an understanding of the world as just, then one of the most vulnerable beliefs to such a bias would be the belief that people are not ultimately in control of their lives. People might convince themselves, as Nietzsche’s priests convinced the masses, that free will exists and therefore suffering is just.

The only remaining question is how belief in a just world *evolved*. I know of no literature on this subject. The just world phenomenon may have evolved, as reciprocal altruism did, to invite mutually beneficial cooperation. One who failed to believe that the

world was just would be less likely to engage in earlier systems of justice—and suffered the resulting cost in fitness. The just world phenomenon is also related to the positive illusions to the extent that an *observer* would want the world to be just. So, to this extent, one might subject the just world phenomenon to the same false advertising and self-deception analysis as the other positive illusions. But the just world effect is primarily about good things happening, not to the self, but to others. As such it may be a non-adaptive over-generalization of the positive illusions into domains where they no longer apply—an exaptation (Gould & Lewontin 1979).

V. CONCLUSION

This article has shown how at least *fifteen* cognitive biases would incline one to believe that free will exists. First, the introduction (I) noted a silence in the philosophical literature about human rationality. That section then distinguished between two important kinds of control and suggested that people might regard themselves as have more control over their lives than they do in fact. The introduction noted that the literature on cognitive biases and heuristics would provide evidence for whether people do have these illusory feelings of control. Finally, the first section introduced a framework for analyzing how cognitive biases can evolve. In part two (II), the article considered Thomas Nagel's problem of *responsibility* and explored how certain cognitive biases might create that problem. In particular, the section argued that four cognitive biases (the endowment effect, outcome bias, anthropomorphic bias, and the FAE) combine to dull the View from Nowhere's paralyzing string—allowing people to jump into the View from Somewhere. The section also explored the evolutionary origin of

these biases in accordance with the heuristics and biases literature and error management theory. In part three (III), the article considered Thomas Nagel's problem of *autonomy* and explored how that problem relates to certain other cognitive biases. The section argued that positive illusions about the self and control create the problem of autonomy by inflating one's sense of control and unpredictability. The section further introduced a novel explanation for how these positive illusions evolved as false advertising and self-deception. Finally, in part four (IV) the article consider some remaining positive illusion of irrational optimism and other cognitive biases which cause belief inertia. In particular, that section considered Nietzsche's error theory of belief in free will, in the larger context of his explanation for how the masses embrace the ascetic ideal, and showed that evidence of the just world phenomenon supports this theory. The section concluded by suggesting that irrational optimism and the just world phenomenon evolved as false advertising, like the other positive illusions, or as a non-adaptive exaptation.

The discovery of so many cognitive biases in favor of belief in free will suggests one important conclusion of this article: if such belief is irrational, there is no one cognitive flaw which explains the phenomenon. In contrast, there is a potent mix of different biases which interact with each other; this combination creates a formidable opponent to rational humility in the face of our human limitations. One consequence of this finding is that it will be difficult to question *all* of the evidence cited in support of this article's conclusions. While it is reasonable to suppose that researchers have exaggerated a few of the biases cited in this article, it is less reasonable to doubt that all of them exist and have no relevance to the free will problem. A second important discovery is that there are no apparent cognitive biases in favor of disbelief in free will.

Non-realists about free will must be motivated by especially cogent arguments in order for their beliefs to survive such cognitive biases—they were probably motivated by appreciation of the biases themselves. Finally, a third important conclusion is that cognitive biases can affect human thinking about this dispute in both subtle and obvious ways. For example, the relevance of the valence effect and the just world phenomenon to the free will problem is *obvious*. It is my hope, however, to have shown how other cognitive biases, such as the endowment effect and anthropomorphic biases, might have more subtle and profound implications for this ancient dispute. These subtle implications are perhaps the most persuasive and important contributions I have to make to this philosophical subject.

These findings are united by at least two themes. First, once one frames the free will problem in terms of actual control and novelist* control, one finds that the evidence for these cognitive biases supports the notion that people regard themselves having more control over their lives than is actual—having control over their lives which approaches novelist* control. Our evolved psychologies may find the prospect of obtaining such novelist* control over our lives to be *seductive*—despite its being impossible. Second, there is a thread running between the assumptions people make about others (part **II**), our inflated sense of control (part **III**), and the importance of happiness to human life (part **IV**). As Gilbert and Malone observe, “dispositional inferences afford the observer a culturally acceptable way of gaining a sense of control over her or his environment, and feelings of control, however illusory, may ultimately yield greater psychological benefits than would logically impeccable inferences” (1995).

In concluding, I must also make some important caveats. First, this article has not concluded that human beings are always irrational. Humans respond to the world in a rational way throughout much of their lives. But their rationality is not perfect. This article has attempted to show where a multitude of cognitive biases converge to put extraordinary pressure upon the human mind. This pressure tests the limits of human rationality. In the absence of such pressure, the mental resources given to the human mind may suffice.

Nor does this article conclude that humans have any control over their lives in any sense. Taylor and Brown make this point in answering Colvin and Block: “[t]he important issue is not whether people believe they can control things they cannot control... but rather whether people believe they can control things more than is actually the case” (1994). The article grants that people have actual control over their lives. This distinguishes normal people from rocks or kleptomaniacs. But the article attempts to show that people regard themselves as having more control than is actual—having control that approach novelist* control. Perhaps we are born “metaphysical megalomania[cs]” (Fischer 2006b).

Correcting this inflated notion of control may have its costs. For example, just as Fischer suggests that non-realists about free will commit a sort of “metaphysical megalomania,” (2006b) so too does he accuse them of a sort of “metaphysical depression” (Fischer 2004a). Indeed, Fischer suggests that non-realists may be vulnerable to *anchoring bias* (Tversky & Kahneman 1974), whereby they rely too heavily upon one trait or piece of information when making decisions. By focusing upon responsibility undermining features of a given situation, at the expense of responsibility

enabling features, the non-realist about free will may reach an overly pessimistic conclusion.

One potential irony of Fischer's claim that non-realists suffer from "metaphysical depression" is that depressed persons show remarkable immunity to many cognitive biases which are relevant to the free will problem. For example, depressed persons are less vulnerable to the fundamental attribution error (Andrews 2001). Similarly, depressed persons seem more immune to all three of the positive illusions identified by Taylor and Brown (1988). They note that more or less depressed persons "are more balanced in self-perceptions," "appear to be less vulnerable to the illusion of control," and "entertain more balanced assessments of their likely future circumstances."

One cause of this depression may be the loss of a chief source of happiness—friendship. If the reactive attitudes reflect faster and cheaper, but less accurate, domain specific neural circuits, and both social competence and the free will illusion depend upon these circuits, then seeing through this illusion may also alienate one from society. There is some anecdotal evidence in support of this hypothesis. For example, Watson observes that "in the same place Einstein speaks of himself as 'a lone traveler,'" he also claims to "not at all believe in human freedom in the philosophical sense" (1987). I have also noted that Russell, another famous non-realist about free will, felt the same isolation:

"Underlying all occupations and all pleasures I have felt since early youth the pain of solitude. I have escaped it most nearly in moments of love, yet even there, on reflection, I have found that the escape depended partly upon illusion. I have known no woman to whom the claims of intellect were as absolute as they are to me, and wherever intellect intervened, I have found that the sympathy I sought in love was apt to fail. What Spinoza call "the intellectual love of God" has seemed to me the best thing to live by, but I have not had even the somewhat abstract God that Spinoza allowed himself to whom to attach my intellectual love. I have loved a ghost, and in loving a ghost my inmost self has itself become spectral. I

have therefore buried it deeper and deeper beneath layers of cheerfulness, affection, and joy of life. But my most profound feelings have remained always solitary and have found in human things no companionship. The sea, the stars, the night wind in waste places, mean more to me than even the human beings I love best, and I am conscious that human affection is to me at bottom an attempt to escape from the vain search for God” (1967-9: 38).

Just as correcting the inflated notion of control may have costs in terms of happiness and social competence, so too may it have costs in terms of precious beliefs. It is telling that Einstein and Russell disbelieved, not just in free will, but also in indeterminism, an afterlife, a personal God, and the efficacy of prayer. In this respect, Einstein and Russell only continue a historical association between determinism and irreligion. For example, Paul Russell cites observations of the same association in Hume’s generation:

“Beattie observes that the doctrine of necessity would be fatal to his religion and moral principles but allows that it may not have the same effect on every other person. Nevertheless, it is, he says, “remarkable, that some of its most distinguished advocates, of whom I shall mention Spinoza, Hobbes, Collins, Hume and Voltaire, were enemies to our faith; whereas of the modern defenders of liberty I do not recollect one who was not a Christian.” Dugald Stewart is also concerned with this connection between the doctrine of necessity and irreligion. He suggests that “it will not be denied, that in the History of Modern Philosophy, the schemes of Atheism and Necessity have hitherto, always been connected together.” (forthcoming)

The suggestion is that whatever alienated these Einstein and Russell from their fellow men also alienated them from God. Indeed, others have noted how both religion (Guthrie 1993) and moral realism (Greene 2002) may play upon cognitive biases in the human psyche. My guess is that the number of cognitive biases favoring belief is even greater, and so the pressure upon human rationality even stronger, in the context of the free will problem than in other contexts. The relevance to religion of the belief inertia

biases, the valence effect, and the just world phenomenon is obvious; similarly, religion may be, as Guthrie emphasizes, even more important to explaining religious beliefs than explaining belief in free will. But many other biases, such as the FAE and the illusion of control, do not seem relevant to religion. This is consistent with Sommers' observation that even some philosophers who have abandoned belief in God and moral realism nevertheless cling to belief in free will (forthcoming).

Although correcting the inflated notion of control will have costs, it will also have benefits. The non-philosophers who have studied these biases have commented upon their importance to questions of public policy. For example, Baron and Hershey remarked as follows:

“Ordinarily, it is relatively harmless to overgeneralize the heuristic of evaluating decisions according to their outcomes. However, when severe punishments (as in malpractice suits) or consequential decisions (as in elections) are contingent on a judgment of poor decision making, insight into the possibility of overgeneralization may be warranted” (1988).

Similarly, Gilbert and Malone warn that:

“In the past year, 1,000 people who thought they knew their acquaintances have been raped by them, 10,000 people who thought they knew their mates have divorced them, and 100,000 people who thought they knew their sovereigns have died as pawns in their wars. Just how capably do we navigate our social worlds? Just how accurate are our understandings of those around us? We do not know. Nobody does. But before we accept the stale contention that people do just fine when psychologists are not manipulating and measuring them, we should probably look around” (1995).

Gilbert and Malone note that one might be blind to guilt in our presence. This article suggests that one may also be blind, in different contexts, not to guilt, but to innocence. In 2005, the United States held 2,186,230 citizens not in quarantine, not in hospitals, but in state or federal prisons.^{viii} Similarly, on January 29, 2002 the President

of the United States declared three foreign countries to be, not states with different values than those of his own country, but an “axis of evil.”^{ix} Some already consider the United States to be in a global World War 3 and encourage the President to “finish off the ‘axis of evil’”.^x Furthermore, scholars have recognized that the proliferation of weapons of mass destruction threatens to destroy humanity (e.g. Rees 2003). In light of these developments, one must note the tragic costs of demonization:

“At the same time, demonizing results in a preoccupation with making the perpetrator suffer (Baumeister, 1997; White, 1995), particularly if demonizing dehumanizes perpetrators (Bandura, Underwood, & Fromson, 1975; Opatow, 1990). The focus on punishing in turn draws attention away from other potentially effective courses of corrective action (Drabek & Quarantelli, 1964)...

Analyses of the origins of evil acts identify vilification and dehumanization of others as one of the enabling conditions for evil (Bandura; [sic] 1999; Baumeister, 1997; Kelman, 1973; Staub, 1999). This gives rise to the *paradox of demonizing*—that those given to perceiving evil in others may be at increased risk for committing evil themselves” (Ellard et al. 2002).

Considering the cognitive biases involved in blaming others (discussed in part II), and the tragic costs associated with these biases, the suffering of our fellow citizens—and perhaps the future of our species—hinges upon our ability to identify and correct the flaws in our evolved psychologies.

One way to fight these cognitive biases is to improve the cognitive capacity and thinking dispositions of people. Such improvements will help to the extent that moral problems arise from conflict between comprehensive reflection in the cortex and cheaper, faster computation in domain specific mental circuits. For example, in the context of the debate between consequentialist and deontological meta-ethics, Joshua Greene reports that subjects who give deontological answers give them much faster than those who give

consequentialist ones (forthcoming). Similarly, Stanovich and West attempted to explain individual differences in rational thought and found that vulnerability to cognitive biases varied with cognitive ability composite and a thinking disposition composite (1998). People with more intelligence and cognitive flexibility were also more rational. This suggests that one way to correct the flaws in our evolved psychologies is just to improve our cognitive capacities and thinking dispositions. If the more cognitive parts of our brain are battling the lower parts of our brain, perhaps we should give the cognitive parts more ammunition.

Finally, lest I succumb to confirmation bias myself, and to distinguish this article from other philosophy articles, I will make some immodest predictions which future research can *test*. These experiments can contrast how those who believe in free will and those who do not perform on various tasks. One's confidence in free will may vary along a wide spectrum, however, and the number of individuals who deny the existence of free will is small relative to the general population. One way to overcome this problem would be to measure, along this spectrum, how much confidence subjects have in free will. One should obtain a correlation between this confidence and expression of the various cognitive biases discussed in this article. For example, if compatibilists and libertarians alike have been accused of defending homunculus theories of human agency, and these theories are expression of anthropomorphic bias, then one should find that non-realists about free will are express this bias less. Similarly, if belief in free will represents an inflated sense of control then one should find that non-realists about free will are less vulnerable to this illusion. If free will represents both a desire for control and a desire for unpredictability but only the former is projected onto others, then one should be able to

identify asymmetries in the philosophical literature such that people describe free will-as-control in the context of others and free will-as-unpredictability in the context of the self. If the just world phenomenon helps motivate belief in free will then people who observe others suffering should not only associate this suffering with unrelated prior bad acts but should also project control onto victims in proportion with their suffering. As a final example, if Nietzsche is correct in supposing that the masses believe in free will because it justifies punishment, and this justification renders their suffering meaningful, then one should find that non-realists about free will tend not to express the just world phenomenon. The same may apply, to lesser or greater degrees, to the other cognitive biases discussed here.

Similarly, if the distorting influence of affect tends to explain compatibilism (and libertarians is not a viable alternative), one would expect non-realists about free will to be less vulnerable to this distorting influence (Nichols & Knobe, forthcoming). If perceivers are “inclined [to] err on the side of assuming immorality regardless of mitigating circumstances” (Haselton & Nettle, forthcoming) then non-realists about free will should tend not to make this assumption. If Joshua Greene’s explanation for deontological views in meta-ethics also explains belief in free will, then people who give more negative answers to questions about whether someone has free will should give them slower than those who give more positive answers (Greene, forthcoming). If Watson is correct to note that diminished appreciation of the reactive attitudes is conducive to disbelief in free will, then one would expect non-realists about free will to express these attitudes less (Watson 1987). Like Einstein and Russell, non-realists may also be more asocial and isolated than others. If people regard themselves and others as possessing control over

their lives which is greater than actual and approaches novelist* control, then surveys should show that people find the notion of someone else having designed their entire life counter-intuitive and *disturbing*. Such a result would show that most prominent compatibilist views—including those of Frankfurt (2002), Watson (1999), Fischer (2004b), Dennett (2003) and Mele (1995: 189-191)—are fundamentally at odds with folk intuitions on the subject because each of these views claims that moral responsibility can survive the prospect of such design. The number of cognitive biases favoring belief in free will also suggests the following double standard. One should regard folk surveys affirming the existence of free will with a healthy skepticism, because they may reflect underlying cognitive biases (such as the fundamental attribution error, the illusion of control, or the just world phenomenon). But one should regard folk surveys denying the existence of free will as especially telling because there are so many cognitive biases which make it difficult to obtain such findings and few, if any, cognitive biases which would make obtaining them easier. Finally, if a multitude of cognitive biases converge to place extraordinary pressure upon people’s rationality when evaluating how much control they and others have, and this pressure tends to explain belief in free will, then the number of known cognitive biases favoring belief in free will should continue to overwhelm the number of known biases, if any, which favor disbelief in free will.

In conclusion, these findings suggest a common usage for the term “free will.” The ancient controversy, according to one characterization, just involves the question of what the term “free will” means. Compatibilists favor a rational conception which fits with the actual control humans have over their lives. The discovery of at least fifteen cognitive biases favoring belief in free will suggests that such a conception of free will

may be too modest. Instead, the common usage of the term “free will” may imply control over one’s life that is greater than actual and approaches novelist* control.

In bridging the gap between the psychological and philosophical literatures, it is also worth noting that, according to an informal poll reported in *Lancet Neurology*, evolutionary psychologists are especially likely to disbelieve in the existence of free will (McCrone 2003). In particular, evolutionary psychologists were more likely to disbelieve than psychiatrists were. This is telling because psychology, including the behaviorist tradition following Skinner (1971) and the psychoanalytic tradition following Freud (1959) and Menninger (1937), has been largely critical of belief in free will. The *Lancet* poll suggests that, on the subject of free will, evolutionary psychologists would tend to be even more radical.

Indeed, this article has shown how at least fifteen cognitive biases favor belief in free will. These biases may create Nagel’s responsibility and autonomy problems, respectively, by inclining humans to feel that they have something approaching novelist* control over their lives. Although the article does not conclude that people are always irrational or have no control over their lives, it does suggest that these biases converge to place extraordinary pressure upon people’s rationality when evaluating how free they and others are. Revising our moral practices in the face of these discoveries, if necessary, may have both costs—in terms of happiness and social competence—as well as benefits—in terms of compassion and public policy. Finally, the article suggests a large number of predictions which future research can test. Until the time of such falsification, these findings suggest that the path to rational belief in free will, if any, is treacherous.

REFERENCES

- Abrams, M. H. (1993). *A Glossary of Literary Terms*, 6th edition. Fort Worth: Harcourt Brace College Publishers.
- Alicke, M. (2006). Blaming badly. Forthcoming in *The Journal of Cognition and Culture*.
- Allport, F. H. (1940). Polls and the science of public opinion. *Public Opinion Quarterly*, 4, 249-257.
- Anderson, C. A., Lepper, M. R., & Ross, L. (1980). Perseverance of social theories: The role of explanation in the persistence of discredited information. *Journal of Personality and Social Psychology*, 39, 1037-1049. (1981-25784-001)
- Andrews, P. W. (2001). The psychology of social chess and the evolution of attribution mechanisms: Explaining the fundamental attribution error. *Evolution and Human Behavior*, 22, 11-29.
- Atran, S. & Norenzayan, A. (2004). Religion's evolutionary landscape: Counterintuition, commitment, compassion, communion. *Behavior and Brain Sciences*, 27 (6), 713-730.
- Bacon, F. (1960). *The New Organon and Related Writings*, edited by Fulton H. Anderson. New York: Liberal Arts Press.
- Baron, J. & Hershey, J. (1988). Outcome bias in decision evaluation. *Journal of Personality and Social Psychology*, 32, 311-328.
- Barrett, J. L., & Keil, F. C. (1996). Conceptualizing a non-natural entity: Anthropomorphism in God concepts. *Cognitive Psychology*, 31, 219-247.
- Baumeister, R. F., Stillwell, A. & Wotman, S. R. (1990). Victim and perpetrator accounts of interpersonal conflict: Autobiographical narratives about anger. *Journal of Personality and Social Psychology*, 59, 994-1005.
- Baumesiter, R. & Vohs, K. (2005). Four roots of evil. In A. Miller (Ed.), *The Social Psychology of Good and Evil* (pp. 85-101). New York: Guilford Press.
- Bleske, A. L., & Shackelford, T. K. (2001). Poaching, promiscuity, and deceit: Combating mating rivalry in same-sex friendships. *Personal Relationships*, 8, 407-424.
- Bornstein, R. F. (1989). Exposure and affect: overview and meta-analysis of research, 1968–1987. *Psychological Bulletin*, 106(2), 265-289.
- Bostrom, N., Ord, T. (2006). The reversal test: eliminating status quo bias in applied ethics. *Ethics*, 116, 656–679.

Brown, J. D. (1986). Evaluations of self and others: Self-enhancement biases in social judgments. *Social Cognition*, 4, 353-376.

Buller, D. J. (2005). *Adapting Minds: Evolutionary Psychology and the Persistent Quest for Human Nature*. Cambridge, MA: MIT Press.

Buss, D. (2000). The evolution of happiness. *American Psychologist*, 55(1): 15-23.

Buss, D. (2003). *The Evolution of Desire: Strategies of Human Mating*. NY: Basic Books.

Callan, M. J., Ellard, J. H., & Nicol, J. E. The belief in a just world and immanent justice reasoning in adults. Forthcoming in *Personality and Social Psychology Bulletin*.

Caporael, L. A. (1986). Anthropomorphism and mechanomorphism: Two faces of the human machine. *Computers in Human Behavior*, 2(3), 215-234.

Carter, R. (1959). Bandwagon and sandbagging effects: Some measures of dissonance reduction. *Public Opinion Quarterly*, 23, 279-287.

Chaiken, S., & Trope, Y. (1999). *Dual process theories in social psychology*. New York: Guilford.

Chapman, G.B. & Johnson, E.J. (2002). Incorporating the irrelevant: Anchors in judgments of belief and value. In T. Gilovich, D. W. Griffin, & D. Kahneman (Eds.), *Heuristics and Biases: The Psychology of Intuitive Judgment* (pp. 120-138). New York: Cambridge University Press.

Chen, M. K., Lakshminarayanan, V., & Santos, L. R. (2006). How basic are behavioral biases? Evidence from capuchin monkey trading behavior. *Journal of Political Economy*, 114(3), 517-537.

Claxton, G. (2000). Whodunnit? Unpicking the 'seems' of free will. In A. Freeman, K. Sutherland, & B. Libet (Eds.) *The Volitional Brain: Towards a Neuroscience of Free Will* (pp. 99-114). Thorverton: Imprint Academic.

Colvin, C. R., & Block, J. (1994). Do positive illusions foster mental health? An examination of the Taylor and Brown formulation. *Psychological Bulletin*, 116, 3-20.

Crocker, J. (1982). Biased questions in judgment of covariation studies. *Personality and Social Psychology Bulletin*, 8, 214-220.

Darwin, C. (1859). *On the Origin of Species by Means of Natural Selection*. London: John Murray.

- Darwin, C. (1871). *The Descent of Man and Selection in Relation to Sex*. New York: Appleton.
- Darwin, C. (1987). *Charles Darwin's Notebooks, 1836-1844*, edited by P. Barrett, P. Gautry, S. Herbert, D. Kohn and S. Smith. Ithaca, New York: Cornell University Press.
- Dawkins, R. (1976). *The Selfish Gene*. New York: Oxford University Press.
- Dennett, D. C. (2003). *Freedom Evolves*. London: Allen Lane; New York: Viking Press.
- Dennis, W. (1953). Animistic Thinking among College and University Students. *The Scientific Monthly*, 76 (4), 247-249.
- Diener, E., Suh, E. M., Lucas, R. E., & Smith, H. L. (1999). Subjective well-being: Three decades of progress. *Psychological Bulletin*, 125, 276-302.
- Doris, J., Knobe J., & Woolfolk, R. Variantism about responsibility. Forthcoming.
- Double, R. (1993). *The Non-Reality of Free Will*. New York: Oxford University Press.
- Einstein, A. (1994). My credo. In M. White & J. Gribbin (Eds.), *Einstein: A Life in Science*. New York: Penguin Books USA Inc.
- Ellard, J. H., Miller, C. D., Baume, T., & Olson, J. M. (2002). Just world processes in demonizing. In M. Ross & D. T. Miller (Eds.), *The justice motive in everyday life* (pp. 350–362). New York: Cambridge University Press.
- Ellsberg, D. (1961). Risk, ambiguity, and the savage axioms. *Quarterly Journal of Economics*, 75, 643–699.
- Evans, J. St. B. T., Barston, J.L., & Pollard, P. (1983). On the conflict between logic and belief in syllogistic reasoning. *Memory and Cognition*, 11, 295-306.
- Fechner, G. T. (1876). *Vorschule der Aesthetik*. Leipzig, Germany: Breitkoff & Hartel.
- Fischer, J. M. (2006a). Introduction: A framework for moral responsibility. In J. Fischer (Ed.), *My Way: Essays on Moral Responsibility* (pp. 1-37). Oxford: Oxford University Press.
- Fischer, J. M. (2006b). The cards that are dealt you. *The Journal of Ethics*, 10, 107–129.
- Fischer, J. M. (2005). Free will, death, and immortality: The role of narrative. *Philosophical Papers*, 34, 379-404.
- Fischer. (2004a). The transfer of nonresponsibility. In J. K. Campbell, M. O'Rourke, and D. Shier (Eds.). *Freedom and Determinism* (at p. 206). Cambridge: MIT Press.

- Fischer, J. M. (2004b). Responsibility and manipulation. *The Journal of Ethics*, 8 (2), 145-177.
- Frankfurt, Harry. (2002). Reply to John Martin Fischer. In S. Buss & L. Overton (Eds.) *Contours of Agency: Essays on Themes from Harry Frankfurt* (pp. 27-32). Cambridge: MIT Press.
- Freud, S. (1959). The 'uncanny'. In J. Riviere (Trans.), *Sigmund Freud: Collected Papers (Vol. 4)* (pp. 368-407). New York: Basic Books, Inc.
- Freud, S. (1989). *The Future of an Illusion*, translated by James Strachey. New York: W.W. Norton & Company.
- Furnham, A. (2003). Belief in a just world: Research progress over the past decade. *Personality and Individual Differences*, 34(5), 795-817.
- Gigerenzer, G. (1996). On narrow norms and vague heuristics: A rebuttal to Kahneman and Tversky. *Psychological Review*, 103, 592-596.
- Gilbert, D. T., & Malone, P. S. (1995). The correspondence bias. *Psychological Bulletin*, 117, 21-38.
- Gintis, H. The evolution of private property. Forthcoming in *Journal of Economic Behavior and Organization*.
- Glynn, J. C., Hayes, F. A. & Shanahan, J. (1997). Perceived support for one's opinions and willingness to speak out: A meta-analysis of survey studies on the 'spiral of silence'. *Public Opinion Quarterly*, 61(3), 452-463.
- Gould, S. J., and Lewontin, R. (1979). The spandrels of San Marco and the Panglossian paradigm: A critique of the adaptationist programme. *Proceedings of the Royal Society of London, Series B*, 205(1161), 581-598.
- Greene, J. D. (2002). The terrible, horrible, no good, very bad truth about morality and what to do about it. Dissertation, Princeton University.
- Greene, J. D. & Cohen J. D. (2004) For the law, neuroscience changes nothing and everything. *Philosophical Transactions of the Royal Society of London B*, 359, 1775-1785.
- Greene, J. D. The secret joke of Kant's soul. Forthcoming in W. Sinnott-Armstrong (Ed.), *Moral Psychology, Vol. 3: The Neuroscience of Morality*. Cambridge: MIT Press.
- Greenwald, A. G. (1980). The totalitarian ego: Fabrication and revision of personal history. *American Psychologist*, 35, 603-618

- Guthrie, S. E. (2002). Animal animism: Evolutionary roots of religious cognition. In I. Pyysiainen & V. Anttonen (Eds.), *Current Approaches in the Cognitive Science of Religion* (pp. 38-67). London and New York: Continuum.
- Guthrie, S. E. (1996). Religion: What is it? *Journal for the Scientific Study of Religion*, 35, 412-419.
- Guthrie, S. (1993). *Faces in the Clouds: A New Theory of Religion*. New York: Oxford University Press
- Hafer, C. L. & Bègue, L. (2005). Experimental research on just-world theory: Problems, developments, and future challenges. *Psychological Bulletin*, 131(1), 128-167.
- Haidt, J. (2001). The emotional dog and its rational tail: A social intuitionist approach to moral judgment. *Psychological Review*, 108 (4), 814-834.
- Haselton M. G. & Buss, D. M. (2000). Error management theory: A new perspective on biases in cross-sex mind reading. *Journal of Personality and Social Psychology*, 78, 81-91.
- Haselton, M. G. & Buss, D. M. (2003). Biases in social judgment: Design flaws or design features? In J. Forgas, K. Williams, & B. von Hippel (Eds.), *Responding to the Social World: Implicit and Explicit Processes in Social Judgments and Decisions*. New York: Cambridge.
- Haselton, M. G. & Nettle, D. The paranoid optimist: An integrative evolutionary model of cognitive biases. Forthcoming in *Personality and Social Psychology Review*.
- Hawkins, S. & Hastie, R. (1990). Hindsight: Biased judgments of past events after the outcomes are known. *Psychological Bulletin*, 107 (3), 311-327
- Hershey, J. & Baron, J. (1992). Judgments by outcomes: When is it justified? *Organizational Behavior and Human Decision Processes*, 53, 89-93.
- Huck, S., Kirchsteiger, G., & Oechssler, J. (2005). Learning to like what you have-- Explaining the endowment effect. *Economic Journal*, 115 (505), 689-702.
- Hume, David. (1957). *The natural history of religion*, edited by H. E. Root. Stanford: Stanford University Press.
- Johnson, E. J., Hershey, J., Meszaros, J., & Kunreuther, H. (1993). Framing, probability distortions, and insurance decisions. *Journal of Risk and Uncertainty*, 7, 35-51.
- Jones, E. E. & Davis, K. E. (1965). From acts to dispositions: The attribution process in

person perception. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (vol. 2, pp. 219-266). San Diego, CA: Academic Press.

Jones, E. E. & Harris, V. A. (1967). The attribution of attitudes. *Journal of Experimental Social Psychology*, 3, 1-24.

Leiter, B. (2004). The Hermeneutics of suspicion: Recovering Marx, Nietzsche, and Freud. In B. Leiter (ed.), *The Future for Philosophy* (pp. 74-105). Oxford: Clarendon Press.

Kahneman, D., Knetch, J. L., & Thaler, R. H. (1991). Anomalies: The endowment effect, loss aversion, and status quo bias. *Journal of Economic Perspectives*, 5(1), 193-206.

Kahneman, D. & Tversky, A. (1979). Prospect Theory: An Analysis of Decision under Risk. *Econometrica*, 47, 263–91.

Kammer, D. (1982). Differences in trait ascriptions to self and friend: Unconfounding intensity from variability. *Psychological Reports*, 51, 99-102.

Kane, R. (1985). *Free Will and Values*. Albany: State University of New York Press.

Kane, R. (1996). *The Significance of Free Will*. New York: Oxford University Press.

Kane, R. (1999). On free will, responsibility and indeterminism: Responses to Clarke, Haji, and Mele. *Philosophical Explorations*, 2, 105-121.

Kane, R. (2001). Some neglected pathways in the free will labyrinth. In R. Kane (Ed.), *The Oxford Handbook of Free Will* (pp. 406-440). New York: Oxford University Press.

Keenan, J. P., Gallup, G., Goulet, N., & Kulkarni, M. (1997). Attributions of deception in human mating strategies. *Journal of Social Behavior and Personality*, 12, 45–52.

Kenrick, D. T., & Gutierrez, S. E. (1980). Contrast effects and judgments of physical attractiveness: When beauty becomes a social problem. *Journal of Personality and Social Psychology*, 38, 131-140.

Knobe, J. & Leiter, B. The case for Nietzschean moral psychology. Forthcoming in Brian Leiter and Neil Sinhababu (eds.), *Nietzsche and Morality*. Oxford: Oxford University Press.

Krebs, J. R., & Dawkins, R. (1984). Animal signals: Mind-reading and manipulation. In J. R. Krebs, & N. B. Davies (Eds.), *Behavioral ecology: an evolutionary approach* (2nd ed., pp. 380– 402). Oxford: Blackwell.

Kruglanski, A. W., & Webster, D. M. (1996). Motivated closing of the mind: “Seizing” and “freezing”. *Psychological Review*, 103, 263–283.

- Langer, E. J. (1975). The illusion of control. *Journal of Personality and Social Psychology*, 32, 311-328
- Langer, E. J. & Roth, J. (1975). Heads I win, tails it's chance: The illusion of control as a function of the sequence of outcomes in a purely chance task. *Journal of Personality and Social Psychology*, 32, 951-955.
- Larson, R. (1989). Is feeling "in control" related to happiness in daily life? *Psychological Reports*, 64, 775-784.
- Larwood, L. & Whittaker, W. (1977). Managerial myopia: Self-serving biases in organizational planning. *Journal of Applied Psychology*, 62, 194-198.
- Lee, A. M., & Lee, E. B. (Eds.). (1939). *The fine art of propaganda: A study of Father Caughlin's speeches*. New York: Harcourt Brace.
- Lerner, M. (1980). *The Belief in a Just World*. New York: Plenum Press.
- Lerner, M. (1965). Evaluation of performance as a function of performer's reward and attractiveness. *Journal of Personality and Social Psychology*, 1, 355-360.
- Lerner, M. & Simmons, C. H. (1966). Observer's reaction to the "Innocent Victim": Compassion or rejection? *Journal of Personality and Social Psychology*, 4 (2).
- Lerner, M. (2003). The justice motive: Where social psychologists found it, how they lost it, and why they may not find it again. *Personality and Social Psychology Review*, 7, 388-399.
- Lewinsohn, P. M., Mischel, W., Chaplin, W. & Barton, R. (1980). Social competence and depression: The role of illusory self-perceptions. *Journal of Abnormal Psychology*, 89, 203-212.
- Marsh, C. & O'Brien, J. (1989). Opinion bandwagons in attitudes towards the common market. *Journal of the Market Research Society*, 31(3), 295-305.
- Mather, M., Shafir, E., & Johnson, M. K. (2000). Misremembrance of options past: Source monitoring and choice. *Psychological Science*, 11(2), 132-138.
- McCrone, J. (2003). Free Will. *Lancet Neurology*, 2(2), 66-66.
- Mehrabian, A. (1998). Effects of poll reports on voter preferences. *Journal of Applied Social Psychology*, 28, 2119-2130.
- Mele, A. (1987). *Irrationality: An Essay on Akrasia, Self-Deception, and Self-Control*. New York: Oxford University Press.

- Mele, A. (1995). *Autonomous Agents: From Self-Control to Autonomy*. New York: Oxford University Press.
- Menninger, K. (1930). *The Human Mind*. New York: Alfred A. Knopf.
- Miller, G. (2001). *The Mating Mind: How Sexual Choice Shaped the Evolution of Human Nature*. New York: Doubleday.
- Nagel, T. (1986). *The View from Nowhere*. New York, Oxford: Oxford University Press.
- Nahmias, E., Morris, S., Nadelhoffer, T., and Turner, J. (2004). The Phenomenology of free will. *Journal of Consciousness Studies*, 11(7-8), 162-179.
- Nayakankuppam, D. & Mishra, H. (2005). The Endowment Effect: Rose-Tinted and Dark-Tinted Glasses. *Journal of Consumer Research*, 32, 390–395.
- Nettle, D. & Keenoo, H. (2005). Schizotypy, creativity and mating success in humans. *Proceedings of the Royal Society B*, 273, 611-615.
- Nichols, S. (2004). Folk psychology of free will. *Mind & Language*, 19, 473-502.
- Nichols, S. How can psychology contribute to the free will debate? Draft.
- Nichols, S. Folk intuitions on free will. Forthcoming in *The Journal of Cognition and Culture*.
- Nichols, S. & Knobe, J. Moral responsibility and determinism: The cognitive science of folk intuitions. Forthcoming in *Nous*.
- Nickerson, R. S. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of General Psychology*, 2(2), 175-22.
- Nietzsche, F. (1886). *Jenseits von Gut und Böse [Beyond Good and Evil]*. Leipzig: Naumann.
- Nietzsche, F. (1967). On the genealogy of morals. In W. Kaufmann and R. Hollingdale (Trans.), *On the Genealogy of Morals and Ecce Homo*. New York: Random House.
- Nietzsche, F. (1968). Twilight of the idols. In W. Kaufmann (Trans.), *The Portable Nietzsche*. New York: Viking Press.
- Nietzsche, F. (2001). *The Pre-Platonic Philosophers*. Urbana: University of Illinois Press.

Noelle-Neumann, E. (1974). The spiral of silence: A theory of public opinion. *Journal of Communication*, 24, 43-51.

O'Connor, T. (2005). Freedom with a human face. *Midwest Studies in Philosophy*, 29, 207-227.

Pereboom, D. (2001). *Living Without Free Will*. New York: Cambridge University Press.

Plott, C. R. & Zeiler, K. (2005). The willingness to pay-willingness to accept gap, the "endowment effect," subject misconceptions and experimental procedures for eliciting valuations. *American Economic Review*, 95(3), 530-545.

Reeder, G. D., & Spores, J. M. (1983). The attribution of morality. *Journal of Personality & Social Psychology*, 44, 736-745

Rees, M. (2003). *Our Final Hour: The Threat to Humanity's Survival*. New York: Basic Books.

Rose, H. and Rose, S. (Eds.) (2000). *Alas Poor Darwin: Arguments Against Evolutionary Psychology*. New York: Harmony Books.

Rosenhan, D. & Messick, S. (1966). Affect and expectation. *Journal of Personality and Social Psychology*, 3, 38-44.

Ross, L. (1977) The intuitive psychologist and his shortcomings: Distortion in the attribution process. *Advances in Experimental Social Psychology*, 10, 174-221

Russell, B. (1957). *Why I am Not a Christian and Other Essays on Religion and Related Subjects*. London: George Allen and Unwin; New York: Simon and Schuster.

Russell, B. (1967, 1968, 1969). *The Autobiography of Bertrand Russell*, 3 vols. London: George Allen and Unwin; Boston and Toronto: Little Brown and Company (Vols 1 and 2), New York: Simon and Schuster (Vol. 3).

Russell, P. Free will and irreligion in Hume's Treatise. Forthcoming in D. Ainslie (Ed.), *Hume's Treatise: A Critical Guide*. Cambridge University Press.

Samuelson, W. & Zeckhauser, R. J. (1988). Status quo bias in decision making. *Journal of Risk and Uncertainty*, 1, 7-59.

Searcy, W.A. & Nowicki, S. (2005). *The Evolution of Animal Communication: Reliability and Deception in Signaling Systems (Monographs in Behavior and Ecology)*. Princeton: Princeton University Press.

- Shogren, J. F., Shin, S. Y., Hayes, D. J., & Kliebenstein, J. B. (1994). Resolving differences in willingness to pay and willingness to accept. *The American Economic Review*, 84 (1), 255-270.
- Skinner, B. F. (1971). *Beyond Freedom and Dignity*. New York: Knopf.
- Smilanksy, S. (2003). Compatibilism: The argument from shallowness. *Philosophical Studies*, 115, 257-282.
- Silverman, I. (1964). Self-esteem and differential responsiveness to success and failure. *Journal of Abnormal and Social Psychology*, 69, 115-119.
- Simon, H. A. (1954). Bandwagon and underdog effects and the possibility of election predictions. *Public Opinion Quarterly*, 18, 245-253.
- Sommers, T. The Illusion of Freedom Evolves. Forthcoming in Spurrett, D., Kincaid, H. Ross, D., Stephens. L. (Eds.), *Distributed Cognition and the Will*. Cambridge, MA: MIT Press.
- Spinoza, B. (1985). Ethics. In E. Curley (Trans.), *The Collected Writings of Spinoza*. Princeton: Princeton University Press.
- Stanovich, K. E. & West, R. F. (1998). Individual differences in rational thought. *Journal of Experimental Psychology: General*, 127(2), 161-188.
- Stein, E. (1996). *Without good reason: The rationality debate in philosophy and cognitive science*. Oxford, England: Oxford University Press.
- Strawson, G. (1986). *Freedom and Belief*. Oxford: Clarendon Press.
- Strawson, G. (1994). The Impossibility of moral responsibility. *Philosophical Studies*, 75, 5-24.
- Strawson. (1998, 2004). Free will. In E. Craig (Ed.), *Routledge Encyclopedia of Philosophy*. London: Routledge. Retrieved August 09, 2005, from <<http://www.rep.routledge.com/article/V014SECT1>>.
- Tajfel, H. & Turner, J. C. (1986). The social identity theory of intergroup behavior. In S. Worchel & W. Austin (Eds.), *Psychology of intergroup relations* (pp. 7–24). Chicago: Nelson-Hall.
- Taylor, S. E. & Brown, J. D. (1988). Illusion and well-being: A social psychological perspective on mental health. *Psychological Bulletin*, 103, 193-210.
- Taylor, S. E. & Brown, J. D. (1994). Positive illusions and well-being revisited: Separating fact from fiction. *Psychological Bulletin*, 116, 21-27.

- Thaler, R. (1980). Toward a positive theory of consumer choice. *Journal of Economic Behavior and Organization*, 1, 39-60.
- Thornhill, R., & Alcock, J. (1983). *The evolution of insect mating systems*. Cambridge: Harvard University Press.
- Trivers, R. (2000). The elements of a scientific theory of self-deception. *Annals of the New York Academy of Sciences*, 907, 114-131.
- Trivers, R. (1985). *Social Evolution*. Menlo Park: Benjamin/Cummings.
- Tversky, A. & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, 185, 1124-1131.
- Vargas, M. (2005). The revisionist's guide to responsibility. *Philosophical Studies*, 125, 399-429.
- Walter, H. (2001). *Neurophilosophy of Free Will: From Libertarian Illusions to a Concept of Natural Autonomy*. Cambridge, MA: MIT Press
- Wason, P.C. (1960). On the failure to eliminate hypotheses in a conceptual task. *Quarterly Journal of Experimental Psychology*, 12, 129-140.
- Watson, G. (1986). Skepticism and free will. *Philosophy and Phenomenological Research*, 46(3), 507-22.
- Watson, G. (1987). Responsibility and the limits of evil. In Schoeman, F. (Ed.), *Responsibility, Character, and the Emotions: New Essays in Moral Psychology* (pp. 256-286). Cambridge: Cambridge University Press.
- Watson, G. (1999). Soft libertarianism and hard compatibilism. *Journal of Ethics*, 3(4), 351-365.
- Werking, K. You are the cards that are dealt you. Draft.
- Williams, G. C. (1957). Pleiotropy, natural Selection, and the evolution of senescence. *Evolution*, 11, 398-411.
- Wright, S. (1932). The roles of mutation, inbreeding, crossbreeding and selection in evolution. *Proceedings of the VI International Congress of Genetics*, 1, 356-366.
- Ybarra, O., Chan, E., & Park, D. (2001). Young and old adults' concerns about morality and competence. *Motivation and Emotion*, 25, 85-100.

Zajonc, R. B. (1968). Attitudinal effects of mere exposure. *Journal of Personality and Social Psychology Monographs*, 9, 1-27.

Zakay, D. (1984). The evaluation of managerial decisions' quality by managers. *Acta Psychologica*, 56, 49-57.

ⁱ This distinction may parallel the alleged distinctions between free will and free action, blameworthiness and moral responsibility, and freedom of indifference and freedom of spontaneity.

ⁱⁱ Kane defines free will as the ability to be the originator and sustainer of one's ends and purposes (1996).

ⁱⁱⁱ It is worth noting that a scholar as prominent as Peter van Inwagen claims to not know what the term "moral responsibility" means.

^{iv} http://www.edge.org/q2005/q05_9.html#pinker

^v The fact that they both published this opinion in the same year, after millennia of philosophizing had failed to produce it, seems to be a coincidence.

^{vi} Friedman, D. Economics and Evolutionary Psychology. <

http://www.daviddfriedman.com/Academic/econ_and_evol_psych/economics_and_evol_psych.html>

^{vii} For an antidote to status quo bias, I recommend Nick Bostrom's "reversal test" (2006).

^{viii} Prison and Jail Inmates at Midyear 2005, 5/06. Presents data on prison and jail inmates, collected from National Prisoner Statistics counts and the Census of Jail Inmates 2005. NCJ 213133 <

<http://www.ojp.usdoj.gov/bjs/abstract/pjim05.htm>>

^{ix} State of the Union address, 2002.

^x Schechter, D. "Is It Time For A Third World War?" Published on Monday, July 17, 2006 by CommonDreams.org