EXPLAINING EVIDENCE DENIAL AS MOTIVATED PRAGMATICALLY RATIONAL EPISTEMIC IRRATIONALITY

MICHAEL J. SHAFFER

Abstract: This paper introduces a model for evidence denial that explains this behavior as a manifestation of rationality, and it is based on the contention that social values (measurable as utilities) often underwrite these sorts of responses. Moreover, it contends that the value associated with group membership in particular can override epistemic reason when the expected utility of a belief or belief system is great. It is also true, however, that it appears to be the case that it is still possible for such unreasonable believers to reverse this sort of dogmatism and to change their beliefs in a way that is epistemically rational. The conjecture made here is that we should expect this to happen only when the expected utility of the beliefs in question dips below a threshold where the utility value of continued dogmatism and the associated group membership is no longer sufficient to motivate defusing the counterevidence that tells against such epistemically irrational beliefs.

Keywords: evidence, rationality, pragmatic rationality, epistemic rationality.

Introduction

Traditionally it has been assumed that people who harbor false beliefs can, at least in principle, be rationally persuaded to correct those beliefs through the use of reason. This importantly involves the assumption that presenting such people with evidence that tells against false beliefs they harbor will bring this about, provided they cooperate and want to have (by and large) true beliefs.[1] In other words, the traditional view is that

---

[1] For critical discussion of the orthodox notion of rationality see Brown (1988), Cohen (1981 and 1982), Stein (1996), Cherniak (1986), Stich (1990), Piatelli-Palmarini (1994), Kahneman, Slovic, and Tversky (1982), Kahneman and Tversky (1996), Gigerenzer (2000), and Manktelow (2012). The problem discussed here involves the idea that the principles of rationality are normative and that they specify how we ought to reason. This, of course, then involves the possibility that we might *choose* to reason in ways that violate those principles. But this, in turn, involves a slight complication as the issue of voluntarism about commitment (for example, belief, acceptance, and so on) formation, since this is a controversial matter. Nevertheless, for practical reasons, the main points made here are approached in terms of the assumption of doxastic voluntarism, although the same point can be made in a bit more complicated way with respect to an involuntarist notion of rational commitment. Concerning such voluntarism/involuntarism about propositional commitments see Alston (1988), Cohen (1992), Steup (1986 and 2000), and Shaffer (2013 and 2017).

epistemic agents are rational and reflectively responsive to evidence. It has recently become increasingly clear, however, that this may not be generally true, and this skeptical conclusion is supported by a growing number of important empirical results, including the selection task, belief perseverance, and conjunction fallacy experiments, among others.[2] These experiments collectively suggest that believers often willfully protect their beliefs in some way that is intentionally motivated by their practical/pragmatic interests and that they are able to undermine counterevidence that tells against false beliefs—often manifestly false beliefs—and inadequately supported beliefs. We see this sort of behavior beyond the lab all too often in both the public and the private spheres, and it should trouble us all, as it appears (at least prima facie) to be profoundly irrational and potentially harmful behavior.

This phenomenon has come to be generally known as motivated irrationality, and evidence denial is just a more specific form of such behavior. I suggest in this paper a model for evidence denial that explains this behavior as a manifestation of rationality, and it is based on the contention that social values (measurable as utilities) often underwrite these sorts of responses. Moreover, I contend that the value associated with group membership in particular can override epistemic reason when the expected utility of a belief or belief system is great. It is also true, however, that it appears to be the case that it is still possible for such unreasonable believers to reverse this sort of dogmatism and change their beliefs in a way that is epistemically rational.[3] The conjecture made here is that we should expect this to happen only when the expected utility of the beliefs in question dips below a threshold where the utility value of continued dogmatism and the associated group membership is no longer sufficient to motivate defusing the counterevidence that tells against such epistemically irrational beliefs. On this basis I suggest that our interventions and attempts to affect such belief changes must take the form of pointing out or creating negative social consequences for such continued commitment. It is important for us to undertake this sort of corrective instruction because these sorts of personal commitments can be publicly and personally harmful. Interestingly, we shall also see that this problem is very closely related to the infamous Quine-Duhem thesis.

---

[2] See Mantelow (2012) for a thorough summary of these results. Of particular interest here are the experiments related to evidence denial and belief perseverance as illustrated in Anderson, Lepper, and Ross (1980), Lord, Ross, and Lepper (1979) and Ross, Lepper, and Hubbard (1975). In a historical sense this phenomenon is interestingly discussed by Peirce (1992) under the rubric of the method of tenacity.

[3] This is issue is a matter of some controversy, especially as it pertains to the backfire effect. See, for example, Nyham and Reifler (2010), Druckamn (2012), Wood and Porter (forthcoming), Rich and Zaragosa (2016), and Rich et al. (2017).

Three Cases of Motivated Irrationality

To make this discussion more concrete, let me introduce three hypothetical but all too real examples to illustrate the phenomenon of evidence denial.[4]

> *Joe Devout:* Joe is a generally rational fellow reasoning under normal conditions who is committed to the existence of the Christian God, understood to be omnipotent, omniscient, and omnibenevolent. Joe attaches great personal value to this commitment and to his membership in a community of people who are like-minded. His life is, in fact, importantly shaped around this commitment. One day, however, Joe notices that his conception of God is incompatible with the existence of evil, and he is clearly aware that there is a nearly limitless set of seemingly undeniably evil events that have occurred in the world. This is a problem because if God is omniscient and omnipotent, then he knows of all these evils, and since he is omnibenevolent he should prevent them. So, Joe learns of a considerable body of evidence against the existence of such a being. But despite learning this he refuses to give up his ideological commitment, due to its great personal value, and to accomplish this refusal he asserts that there is in fact no evil in the world at all. He maintains his degree of commitment to the claim that the Christian God exists at the same degree prior to his learning the new bit of counterevidence.

In this case Joe learns of a body of facts that undermines his commitment to the ideologically grounded belief that the omnipotent, omniscient, and omnibenevolent God of orthodox Christianity exists. He becomes aware of the so-called problem of evil. Rather than acknowledge this evidence and update his beliefs appropriately, however, he uses his ideological commitment to dismiss the counterevidence so as to preserve his valued personal commitment.

Let's consider this second similarly constructed case:

> *Jane Novax:* Jane is a generally rational gal reasoning under normal conditions who is committed to the claim that the standard regimen of childhood vaccinations significantly increases the chance of a child's being autistic. She attaches great personal value to this commitment and to her membership in a community of people who are like-minded. Her life is, in fact, importantly shaped around this commitment. Jane notices, however, that there are no good studies in support of this claim and that there are very many reputable and reliable studies that demonstrate that the claim is almost certainly false. So, she has a considerable body of evidence against the claim that the standard regimen of childhood vaccinations increases the chance of a child's being autistic. But she refuses to give up her ideological commitment to this claim on learning of this counterevidence, due to her commitment's great personal value, and to

---

[4] In these cases the assumption is being made that the relevant commitments involved are non-evidence-based. This is, of course, likely an idealization with respect to real-world cases, but the formal model introduced here nevertheless illuminates the factors involved.

accomplish her refusal she asserts that the studies are phony and are, in fact, part of a conspiracy perpetrated by the pharmaceutical companies to sell their products. So she uses her ideological commitment to establish that this apparent counterevidence is not actually disconfirmatory and maintains her initial degree of commitment to the claim that the standard regimen of childhood vaccinations significantly increases the chance of a child's being autistic.

In this case Jane learns of a body of facts that undermines her commitment to the ideologically adopted claim that the standard regimen of childhood vaccinations is harmful and causally connected to the incidence of autism. Rather than acknowledge that evidence and make appropriate adjustments in her belief state, however, she uses her ideological commitment simply to neutralize the counterevidence by concocting a conspiracy theory.

Finally, consider a third, slightly different sort of case:

Donald Denier: Donald is a generally rational fellow reasoning under normal conditions who is committed to the claim that climate change is not the result of the human use of fossil fuels. He attaches great personal value to this commitment and to his membership in a community of people who are like-minded. His life is, in fact, importantly shaped around this commitment. He notices, however, that there are very few studies in support of this claim and that they are all of questionable quality. Moreover, there are very many reputable and reliable studies that demonstrate that it is overwhelmingly likely that climate change is the result of the human use of fossil fuels. So, Donald has a considerable body of evidence for the claim that climate change is the result of the human use of fossil fuels. But he refuses to give up his ideological commitment to this claim, due to its great personal value, and to accomplish his refusal he asserts that the studies that serve as counterevidence are phony and are, in fact, part of a conspiracy perpetrated by liberals. He adopts the position that the supposed counterevidence is not actually disconfirmatory and increases the degree of his commitment to the claim that that climate change is not the result of the human use of fossil fuels, rather than reducing his degree of confidence in that claim.

Notice that unlike the agents in the Joe Devout and Jane Novax cases, in Donald Denier the agent in question does not simply maintain his level of commitment to the claim in question in response to the counterevidence he becomes aware of. Rather, Donald *increases* his degree of commitment to this claim upon learning the counterevidence. Thus, this case is designed to represent the salient facts involved in cases of what is known as the backfire effect. Not only does Donald ignore evidence on the basis of his ideological commitment, he also strengthens that commitment in response to compelling counterevidence.

So what are we to make of these sorts of cases? Are the agents involved simply stupid, stubborn, or irrational? Are they simply violating

fundamental epistemic rules or is something more pernicious and subtle going on? In order to address and perhaps correct these sorts of behaviors involving such putative irrationality we need first to find a reasonable explanation of just what is going on in these sorts of cases of evidence denial. Let us begin, then, with a brief examination of the concept of rationality and some associated issues involving human cognitive behavior in order to make some sense of these sorts of cases.

Rationality, Belief Revision, and Conditionalization

At least since Aristotle boldly proclaimed that man is a rational animal, philosophers have typically assumed that humans are by and large rational beings, and this is supported by a compelling form of the Davidsonian principle of charity, which asserts that we ought to prefer explanations of human behavior that treat our behaviors as rational, at least when it is possible to do so. Let us call this assumption *the rationality thesis*. Of course, there has been ongoing philosophical debate both about just which principles make up human rationality and about how well we collectively meet those norms of reasoning in our cognitive endeavors. So, let us call the first project—that of determining which principles constitute human rationality—the normative project, and let us call the second project—that of determining how rational humans are relative to some identified standard—the descriptive project. It should be clear, then, that, so understood, the descriptive project depends on the normative project, and that the rationality thesis involves the assumption that human rationality is constituted by some set of principles of good reason that we all exhibit and that we typically conform to, even if we sometimes make errors in reasoning. But this way of looking at the matter immediately invites addressing the normative project. We cannot fairly assess the status of the rationality thesis via confronting the descriptive project without understanding first what constitutes the best normative model of human rationality.

The prevailing—or even orthodox—model of human rationality takes an agent's epistemically rational behavior to involve belief change and creedal states (both synchronic and diachronic) that conform to normative principles governing belief revision and the probabilistic credences agents attach to such beliefs. There are, of course, numerous technical problems involved in both of these key components of the orthodox model, and they are at best complementary, since they cannot easily be integrated, but the model is assumed to be correct in principle. Let us look briefly at the most prominent theory of belief revision. Carlos Alchourrón, Peter Gärdenfors, and David Makinson developed the AGM (Alchourrón-Gärdenfors-Makinson) theory of belief revision in the 1980s, and a number of related theories have arisen as a consequence.[5] These theories are

---

[5] See Alchourrón, Gärdenfors, and Makinson (1985), Gärdenfors (1988), and Levi (1996).

fundamentally based on the concept of a belief state, a belief set, or a corpus of beliefs, K, typically satisfying the following minimal conditions (where it is assumed that belief states are given a representation in some language L):

- (*Df* BS) A set of sentences, K, is a belief state if and only if (i) K is consistent and (ii) K is objectively closed under logical implication.

The content of a belief state is then defined as the set of logical consequences of K. Given this basic form of epistemic representation, the AGM-type theories are intended to be normative theories about how a given belief state that satisfies the definition of a belief state is related to other belief states satisfying that definition relative to: (1) the addition of a new belief b to K_i or (2) the retraction of a belief b from K_i, where b $\in$ K_i. Belief changes of the latter kind are termed *contractions*, but belief changes of the former kind must be further subdivided into those that require giving up some elements of K_i and those that do not. Additions of beliefs that do not require giving up previously held beliefs are termed *expansions*, and those that do are termed *revisions*.[6]

Given AGM-style theories, the supposed a priori grounded principles that govern the rational dynamics of beliefs are simply the epistemically normative rules that govern rational cases of contraction, revision, and expansion of belief states. These rules are the AGM postulates, and the fundamental insight behind these theories is then that belief changes should be fundamentally conservative in nature. In other words, in belief changes one ought to make the minimal alterations necessary to incorporate new information and ought to maintain or restore logical consistency. This fundamental assumption is supposed to be justified a priori in virtue of a principle of informational economy. This principle reflects the idea that information is intrinsically and practically valuable, and so we should retain it at all costs unless we are forced to do otherwise. So, while the details are not important here, the revision operations on belief states are normatively restricted so as to obey a principle of minimal mutilation and a principle of consistency maintenance. But what about the credences that we attach to beliefs in the orthodox account of rationality? Creedal states are not mentioned in AGM models of belief, but they importantly reflect our degrees of confidence with respect to our beliefs. So, fleshing out the orthodox model requires supplementing the belief revision theory with a theory of credences or of probabilistic rationality.

The orthodox model of probabilistic rationality that complements the orthodox model of belief revision is Bayesianism (see Howson and Urbach [1993]). With respect specifically to creedal states (that is, our

---

[6] In point of fact, the AGM theory really holds only that there are two dynamical operations on belief states, because revision is defined in terms of expansion and contraction.

probabilistically graded partial beliefs) Bayesianism specifies well-known rules of probabilistic rationality, and the justification of these principles was traditionally secured in the a priori philosophical way, just as the belief revision postulates. Specifically, our creedal states and their updating are supposed a priori to abide by conditionalization if they are rational. Conditionalization is understood as follows:

- (Conditionalization) $P'(A) = P(A \mid e)$.

So understood, this principle says that upon learning evidence e at t, your new probability function $P'(\cdot)$ at $t + 1$ defined over your beliefs should be equal to the conditional probability of A given e at t. In other words, you should incorporate e into your initial creedal belief state at t and adjust the probabilities defined over your partial beliefs appropriately to yield a new rational probability function at $t + 1$ obeying the axioms of the probability calculus. More specifically, the conditional probability of A given e is defined a priori by Bayes' Theorem as follows:

- (Bayes' Theorem) $P(A|e) = P(A)P(e|A)/P(e)$.

Together conditionalization and Bayes' Theorem yield a potent normative theory of synchronic and diachronic probabilistic rationality. They tell what creedal states we ought to have, given our evidence. So, given the orthodox model of rationality, if we are to be rational, every time a new bit of evidence comes in we are supposed to revise our belief system to incorporate that datum and then adjust our creedal states with respect to our belief states according to Bayesian conditionalization. Now, if we look at the Joe Devout, Jane Novax, and Donald Denier cases through the lens of the orthodox model, it looks clear that Joe, Jane, and Donald are irrational. Each of them appears to violate the normative principles that constitute this model of rationality, and it would seem to be the case that we should just educate them and perhaps present them with further counterevidence in order to effect rational revision of their belief and creedal states. But, as we shall see, things are far more complex and subtle than this.

The Experimental Challenge to the Orthodox Model

To begin to see this, let us first remind ourselves that there is currently a large body of troubling experimental results that appear to have bearing on this particular model of human rationality and on the rationality thesis in general. What is most interesting about this particular issue is that it has rarely been the case that philosophers and psychologists who have been involved as disputants in either or both of the normative and

descriptive debates have challenged the rationality thesis itself. This is in part the result of the lingering essentialism of the Aristotelian version of the rationality thesis. For Aristotle it was a matter of philosophical definition that humans are rational. Man is, *by definition*, a rational animal. In other words, we are essentially rational beings such that if we were to lose our capacity for being rational we would no longer qualify as human except in a purely nominal sense. In fact, Aristotle tells us straightforwardly that a human whose desires and appetites subordinate his or her reason literally becomes a beast and not a human in terms of function. So, according to this view, we are all necessarily rational in principle, and this is because on the traditional view the rationality thesis has been taken to be a purely philosophical, definitional, and a priori matter. As a result, the matter of the truth of the rationality thesis is supposed to be totally insulated from the world of empirical results. If one grants this philosophical assumption, then the normative project is just the project of analyzing the specific content of the concept of human rationality via the a priori methods of philosophy, and it is just a given that all humans are rational.

What is of particularly great interest here, then, is that beginning in the 1960s a number of psychologists began to conduct *experimental* studies of human rationality and in so doing changed the methodological dynamic associated with this issue. Most notorious in this regard were the various studies conducted by Daniel Kahneman and Amos Tversky and by Peter Wason and Philip Johnson-Laird.[7] What these and other psychologists ultimately found in their experiments was, at the time, quite startling. In effect, they found that actual human reasoning apparently fails to conform to the basic capacities that had been assumed to be part of human rational ability in the philosophical sense. This was shown to be true with respect to core aspects of inductive, deductive, and abductive reasoning. But what is most interesting about this story is that a number of these researchers then concluded on the basis of the sum total of the results of experiments of these types that the extent of human rationality is, at very least, far more restricted than philosophical models of rationality had assumed. They did not, then, merely use empirical methods to address the descriptive project, and they were not content to simply claim that the results showed that we do not meet the norms of human rationality in many particular cases of thinking. Rather, what they concluded was the altogether more radical claim that the rationality thesis itself was false, and so concluded that the normative project is itself radically misguided. Essentially their argument amounts to the claim that humans' ability to reason diverges so systematically from any proposed norms offered up in

---

[7] See Kahneman and Tversky (1979), Kahneman and Tversy (1982), Kahneman and Tversky (1996), Wason (1968a, 1968b, 1968c, and 1977), Wason and Johnson-Laird (1970), and Johnson-Laird and Wason (1977).

the philosophical analyses of human rationality that we should simply reject the rationality thesis lock, stock, and barrel. Aristotle was just wrong, philosophers do not know what they are talking about, and we are fundamentally irrational beings.

So the contemporary debate about human rationality crucially turns on a large body of experimental data concerning epistemic performance. On the one hand, it seems that these systematic and widespread failures on the part of humans to meet the philosophically established norms of good reasoning may, when suitably interpreted and correctly marshaled, imply that humans are not as rational as traditionally assumed. On the other hand, perhaps they can be explained away in some manner that preserves the idea that we are, at least mostly, rational. This second line of thought needs to be considered in a bit more detail at this juncture, and it is a desirable alternative in that it preserves the broad assessment of humans as rational and helps us to make sense of their behavior. So in order to make sense of this alternative, it is useful to introduce a particular conceptual distinction. There are two distinct ways in which we can view the sorts of errors that appear to be present in these sorts of experiments, and a useful way to look at them derives from Noam Chomsky's work in linguistics. In explaining how we use language successfully despite the fact that we often make mistakes, Chomsky (1975 and 1986) introduced the distinction between *performance* and *competence*. For example, we are competent speakers of English despite the fact that we often make grammatical errors in our speaking and writing. Competence, then, is a matter of our ideal functioning with respect to linguistic use, whereas performance is a matter of our actual track record in following and/or applying the constitutive principles of a language in practice. Performance errors are specifically supposed to occur when we are tired, when we are distracted, when other cognitive systems interfere with our linguistic processing, and so on. Such errors occur due to psychological and situational interference not under our control. But so long as the errors in question are not systemic and uncorrectable in better conditions, such results need not be taken to negatively impact the claim that humans possess linguistic competence. It is just that sometimes external factors beyond our control adversely impact our linguistic performance.

The structure of the debate about human rationality suggests that we might apply this sort of distinction in the case of the disparity between the normative principles of rationality and our actual reasoning. In essence, we might say that we have a certain sort of rational and ideal competence with respect to the constitutive principles of rationality despite the fact that we sometimes make errors in actual reasoning due to external factors beyond our direct control. Thus one can seemingly preserve the claim that we are rational even in the presence of errors, especially if such errors are not systemic and uncorrectable. But the problem that this paper focuses on arises despite this clever defense of the claim that humans are rational

in light of empirically established errors in reasoning. Specifically, in the context of motivated irrationality things are a bit more complicated, and what we shall find is that it is much more reasonable and constructive to believe that rational commitment involves *nonepistemic factors* that, if ignored, mislead us about the nature of evidence denial. So, really, the problem that must be addressed here concerns what we should say about the cognitive behavior of people who *willfully* reason in ways that violate the principles that are constitutive of rational competence. These people are not seemingly making mistakes because of external factors beyond their control that cause their reasoning to deviate from ideal competence. Rather, they are *choosing* to intentionally violate those rules. Specifically, the problem to be addressed involves a form of what has come to be called *motivated irrationality*, and it is based on the idea that ideologically driven thinkers sometimes consciously and intentionally violate the rules that are constitutive of human rational competence in order to preserve their ideological beliefs. Often they knowingly ignore or defuse counterevidence and counterarguments that negatively implicate their ideological commitments. Sometimes, in fact, the degree of commitment to such ideological beliefs knowingly *increases* in response to such negative evidence. This latter, more specific, phenomenon is known as the backfire effect. What I shall suggest here is, first, that such motivated irrationality is rationally accomplished by appeal to specific kind of value that ideologically driven thinkers attach to ideological commitments and not by appeal to the epistemic status of such commitments. In essence, they defuse counterevidence by appeal to a form of rational decision making. Second, I shall suggest that the proper sorts of interventions that will predictably fix this sort of irrationality do not involve purely or even primarily epistemic interventions. This implies that the orthodox model of rationality is importantly deficient in the sense that it treats the notion of rational commitment as a purely epistemic matter and in doing so would force us to understand evidence denial as a failure of epistemic rationality in a way that fails to capture the most salient feature of that phenomenon (that is, the role of pragmatic considerations in rational commitment).

So, at least from the perspective of the orthodox model of rationality, this sort of relatively common bad epistemic behavior presents us with something of a puzzle. This is because of the following considerations. First, the cognitive agents in these situations appear to be in an epistemic position to believe what the evidence warrants, but they choose to do otherwise, in violation of the orthodox model of rationality. Second, they appear to hold evidence contravening beliefs on the basis of what sometimes appear to be nonepistemic reasons. As a result, what is common to all three cases is the idea that the agents involved value their respective ideological commitments in a nonepistemic and contra-evidential manner—they specifically derive great value from being members of communities of like-minded thinkers despite the counterevidence they possess—and in so doing

are able to resist the proper normative implications of the counterevidence they become aware of. This immediately and unambiguously suggests that those commitments are not purely functions of their epistemic states. Rather, such agents seem to be basing their commitments, at least in part, on important and potent nonepistemic factors. Specifically, they appear to be basing their commitments on the expected utilities of those commitments, specifically expected utilities that are importantly linked to groupthink. Failure to appreciate this take on the matter would force us to conclude that these agents are simply irrational in the sense that they fail to possess rational competence or in the sense that they are merely making performance errors with respect to the orthodox principles of rationality. But as the stories are set up the agents involved are generally rational and not subject to external factors that cause them to make performance errors. So this does not seem to be right, and it would require attributing widespread failures of rational competence and/or performance that are not easily explained and that would violate the principle of charity when it comes to their cognitive behavior.[8] Let us, then, turn our attention to developing a better explanation of these sorts of cases in light of what we have learned about rationality in general and what we have conjectured about cases of evidence denial.

Explaining Evidence Denial and the Quine-Duhem Thesis

In the Joe, Jane, and Donald cases the agents learn of evidence e that bears on an ideological commitment A such that:

- (Counter Evidence) P(A|e) < P(A).

This is just the standard probabilistic notion of disconfirming evidence. As we have already seen, however, in light of this situation the agents ought rationally to conditionalize and accept that e is disconfirming evidence with respect to A so that $P'(A) < P(A)$. But none of the agents in the three cases does this, and one might be tempted to claim that all three agents are just irrational in the sense of failing to possess the proper competence and leave it at that. They all just need to be corrected and learn to properly employ the postulates of AGM belief revision and Bayesian conditionalization. But this is an entirely unnuanced response, and it does not reveal precisely how these agents—each of whom appears to be otherwise rationally competent and free from external factors that might cause performance errors—manage to do this. What will be suggested here is that it is plausible to suppose that the agents in our three cases adopt one of the following principles with respect to e's being disconfirming evidence for A:

---

[8] See Davidson (1973 and 1974) on the problems with doing this.

- (Ideological Immunization—1) (EV(A) > δ) & [(EV(A) > δ) → (A → ¬e)].
- (Ideological Immunization—2) (EV(A) > δ) & {(EV(A) > δ) → ¬[(P(A|e) < P(A)]}.

That is to say that these agents hold that when the expected value of A exceeds some crucial threshold and in virtue of the value that A has, A implies either that e is false or that e does not disconfirm A.[9] In effect, on either one or the other of these bases, they deny that e is really counter-evidence with respect to A because of the value that A has. Alternatively, agents of these sorts may accept that e is disconfirming (but not refuting) evidence for A where A has a high expected value, but upon learning e the agent *increases* his or her valuation of A in order to maintain or even strengthen that commitment. Call this notion Ideological Immunization—3, and it is a more specific version of Ideological Immunization—2. All three of these curious but illuminating strategies identify ideological commitments as something more than epistemic commitments. Specifically, they all involve expected values, and this is the key to the problem of understanding what is going on in the Joe, Jane, and Donald vignettes.

What, then, is the expected value of a proposition? According to standard decision theory, the expected value of a proposition is understood as follows:

- (Expected Value) EV(A) = P(A) × U(A).

This just tells us that the expected value of a proposition is the product of its probability and its subjective utility, where subjective utilities are just reflections of rational personal preferences. The rationality of such preferences is merely the internal coherence of such values, and this is reflected in the axioms of decision-theoretic utility that assure that such preferences are, for example, transitive. So, any claim with a non-zero probability and a nonzero utility of this sort will have a nonzero expected value. But since expected values are mathematical products, a claim with a marginal probability can still have a *very large* expected value when its subjective utility is great. Of course, this is typically true of ideological commitments of the sort at issue here. This realization is crucially important to the matter at hand. It is simply too naïve to claim that the agents in the Joe, Jane, and Donald cases are just irrational in

---

[9] The issue of just what the threshold value is may vary from individual to individual, owing to the fundamental nature of utilities (that is, they are measures of rational subjective preference). But, that said, the reasonable conjecture made here is that the threshold can be found where the value of group membership that is conferred by adoption of the commitment in question is significantly greater than the value of having the relevant true commitment.

the ordinary sense of broad competence or performance. They are possessed of basic logical and probabilistic competence and are not making errors due to external factors like tiredness, and so forth. Their behavior is both more intentional and cleverer than that, and correcting such behavior crucially requires understanding this, especially since all three of the ideological immunization strategies can be repeatedly employed to defuse virtually any counterevidence that could be brought to bear against their commitments simply by ratcheting up their assessments of the utility of those commitments. So, the agents' apparent dogmatism is consciously motivated, sinisterly rational, and often very effective, and we should understand that in cases like these we are really dealing with *motivated pragmatically rational epistemic irrationality*. It is, then, also deeply interesting to note that this sort of epistemically irrational behavior is importantly related to Quine's particular version of the Quine-Duhem thesis. In its simplest guise, Quine's version of the thesis incorporates two components. First, it involves conformational holism. This just means that counterevidence is counterevidence only with respect to a total body of beliefs that are assumed in a conformational/disconfirmational context. Second, it involves the contention that any claim may be retained in light of apparent refutation from counterevidence by making appropriate changes in one's standing body of beliefs (see Quine [1951] and Sanford [2017]). Both of these subtheses are deeply important to the notions of theory confirmation, theory refutation, and rational belief in general, and they are crucial to the sort of evidence denial that is the central topic of this paper. As Quine claims,

> The totality of our so-called knowledge or beliefs, from the most casual matters of geography and history to the profoundest laws of atomic physics or even of pure mathematics and logic, is a man-made fabric which impinges on experience only along the edges. Or, to change the figure, total science is like a field of force whose boundary conditions are experience. A conflict with experience at the periphery occasions readjustments in the interior of the field. But the total field is so underdetermined by its boundary conditions, experience, that there is much latitude of choice as to what statements to reevaluate in the light of any single contrary experience. No particular experiences are linked with any particular statements in the interior of the field, except indirectly through considerations of equilibrium affecting the field as a whole. (1951, 42–43)

Notice, then, that given Quine's version of the Quine-Duhem thesis believers can voluntarily revise their beliefs/commitments to prevent refutation in a variety of ways that need not abide by the orthodox principles of rationality.[10] So, Quine and like-minded thinkers appear to be committed

---

[10] Quine (1995) is clear that one can adjust one's total commitments in light of apparent refutation in any number of ways. Importantly, one can, in principle, reject the counterevidence to defuse any such refutation.

to a form of direct doxastic voluntarism, the view that we have direct voluntary control over what we believe, and they allow, then, for the possibility that we can voluntarily commit/decommit ourselves to propositions and in so doing can defuse apparent refutations of any belief either by contracting with respect to the counterevidence or by revising/expanding our beliefs so as to defuse the refutation (for example, by constructing an appropriate conspiracy theory). So, these defenders of the Quine-Duhem thesis allow that we can at least sometimes efficaciously choose to disbelieve propositions that negatively impact our other beliefs, including counterevidence itself. But this leaves the question of the rationality of such maneuvers unexplained, poorly explained, or perhaps even inexplicable in terms of the orthodox model of rationality. As we have seen, however, this sort of cognitive phenomenon is totally explainable and even predictable in terms of the hybrid model of motivated irrationality suggested here. Something like Quine's own notion of entrenchment is at work in cases of evidence denial when we understand entrenchment as a pragmatically rational phenomenon, and this allows us to see that there is a fully understandable but not entirely orthodox notion of rational commitment involved in denying evidence. In any case, with this explanation in hand we can consider the final piece of the puzzle: How should we critically deal with evidence denial and evidence deniers?

Conclusion: Inflating the Cost of Irrationality

The answer to this question is to be found in the hybrid explanatory model of the Joe, Jane, and Donald cases. First, in virtue of the general moral principle that we have a duty to prevent avoidable and unnecessary harms (that is, the principle of nonmalfeasance), we have an obligation ourselves to try to prevent the sorts of harmful irrationality exhibited in those cases, and the efficacious way to accomplish this good is revealed via the explanations of ideological immunization involved in evidence denial offered earlier. Specifically, it should be clear that the sort of motivated irrationality exhibited by Joe, Jane, and Donald depends deeply on groupthink. That is how the personal value of their commitments is grounded and maintained (via the utility of group membership). In virtue of this value derived from and reinforced by group membership the expected values of certain sorts of ideological commitments are used to protect such commitments from rational refutation or disconfirmation. But, as we have seen, we should not succumb to the temptation to label these sorts of behaviors as blanket indicators of epistemic irrationality with respect to the orthodox model either in the sense of lack of competence or in terms of performance errors. Rather, we should understand that subjects like Joe, Jane, and Donald do satisfy the rationality thesis but exhibit a hybrid form of rationality that in part depends on nonepistemic and distinctly pragmatic factors.

So, it would appear to be the case that it is these latter factors with which we must contend if we are to correct such bad and harmful cognitive behavior. But there is hope here too. As recent studies have shown, agents who exhibit behavior of the sort involved in evidence denial and/or the backfire effect can sometimes be brought back into accord with the canons of epistemic rationality, but not merely by piling on more epistemic criticism or providing epistemic instruction.[11] The conjecture made here is that such corrective alterations in epistemic behavior can be expected to happen only when we can reduce the expected value of an ideological commitment that is used to immunize that commitment from disconfirmation below the threshold that engages the ideological immunization strategy the agent is employing. As has been stressed here, however, it seems that this cannot be achieved by deploying merely epistemic resources and epistemic instruction. Rather, what we need to do instead is to focus specifically on the reduction of the value of those commitments, principally by *inflating the costs of those commitments*. This is likely best achieved by attacking the communities that give rise to the perceived value of those commitments or by showing that there is little value in belonging to such a group. In other words, we must make it *socially undesirable to be members of those communities*, and this suggests that the use of methods such as public ridicule, isolation, deprogramming, public awareness campaigns, and so on will be far more effective in achieving the correction of such motivated irrationality than purely epistemic interventions. Essentially, if we hope to correct these epistemic and moral wrongs, what we must do is reduce the expected values of the ideological commitments of these sorts of people below the tipping point where epistemic reason can take over and their rational epistemic competence is engaged.

Department of Philosophy
St. Cloud State University
CH 365
720 4th Ave. S.
St. Cloud, MN 56301
USA
mjshaffer@stcloudstate.edu

References

Alchourrón, C., P. Gärdenfors, and D. Makinson. 1985. "On the Logic of Theory Change: Partial Meet Functions for Contraction and Revision." *Journal of Symbolic Logic* 50:510–30.

---

[11] See Wood and Porter (forthcoming), Rich and Zaragosa (2016), and Rich, Van Loon, Dunlosky and Zaragosa (2017).

Alston, W. 1988. "The Deontological Conception of Epistemic Justification." *Philosophical Perspectives* 2:257–99.

Anderson, C., M. Lepper, and L. Ross. 1980. "Perseverance of Social Theories: The Role of Explanation in the Persistence of Discredited Information." *Journal of Personality and Social Psychology* 39:1037–49.

Brown, H. 1988. *Rationality*. New York: Routledge.

Cherniak, C. 1986. *Minimal Rationality*. Cambridge, Mass.: MIT Press.

Chomsky, N. 1975. *Reflections on Language*. New York: Random House.

———. 1986. *Knowledge of Language*. Westport, Conn.: Praeger.

Cohen, L. J. 1981. "Can Human Irrationality Be Experimentally Demonstrated?" *Behavioral and Brain Sciences* 4:317–31.

———. 1982. "Are People Programmed to Commit Fallacies?" *Journal for the Theory of Social Behavior* 12:251–74.

———. 1992. *An Essay on Belief and Acceptance*. Oxford: Clarendon Press.

Davidson, D. 1973. "Radical Interpretation." *Dialectica* 27:314–28.

———. 1974. "On the Very Idea of a Conceptual Scheme." *Proceedings and Addresses of the American Philosophical Association* 47:5–20.

Druckamn, J. 2012. "The Politics of Motivation." *Critical Review* 24:199–216.

Gärdenfors, P. 1988. *Knowledge in Flux*. Cambridge, Mass.: MIT Press.

Gigerenzer, G. 2000. *Adaptive Thinking*. Oxford: Oxford University Press.

Howson, C., and P. Urbach. 1993. *Scientific Reasoning: The Bayesian Approach*. 2nd ed. Chicago: Open Court.

Johnson-Laird, P., and P. Wason, eds. 1977. *Thinking*. Cambridge: Cambridge University Press.

Kahneman, D., and A. Tversky. 1979. "On the Interpretation of Intuitive Probability: A Reply to Johnathan Cohen." *Cognition* 7:409–11.

———. 1996. "On the Reality of Cognitive Illusions." *Psychological Review* 103:582–91.

Kahneman, D., P. Slovic, and A. Tversky. 1982. *Judgment Under Uncertainty: Heuristics and Biases*. Cambridge: Cambridge University Press.

Levi, I. 1996. *For the Sake of the Argument: Ramsey Test Conditionals, Inductive Inference, and Nonmonotonic Reasoning*. Cambridge: Cambridge University Press.

Lord, C., L. Ross, and M. Lepper. 1979. "Biased Assimilation and Attitude Polarization: The Effects of Prior Theories on Subsequently Considered Evidence." *Journal of Personality and Social Psychology* 37:2098–2109.

Manktelow, K. 2012. *Thinking and Reasoning*. New York: Psychology Press.

Nyham, B., and R. Reifler. 2010. "When Corrections Fail: The Persistence of Political Misperceptions." *Political Behavior* 32:303–30.

Peirce, C. S. 1992. "The Fixation of Belief." In *The Essential Peirce*, vol. 1, edited by N. Houser and C. Kloesel, 109–23. Bloomington: Indiana University Press.

Piatelli-Palmarini, M. 1994. *Inevitable Illusions*. New York: Wiley.

Quine, W. 1951. "Two Dogmas of Empiricism." In *From a Logical Point of View*, 2nd ed., 20–46. Cambridge, Mass.: Harvard University Press.

———. 1995. *From Stimulus to Science*. Cambridge, Mass.: Harvard University Press.

Rich, P., and M. Zaragosa. 2016. "The Continued Influence of Implied and Explicitly Stated Misinformation in News Reports." *Journal of Experimental Psychology: Learning, Memory and Cognition* 42:62–74.

Rich, P., M. Van Loon, J. Dunlosky, and M. Zaragosa. 2017. "Belief in Corrective Feedback for Commonsense Misconceptions: Implications for Knowledge Revision." *Journal of Experimental Psychology: Learning, Memory and Cognition* 43:492–501.

Ross, L., M. Lepper, and M. Hubbard. 1975. "Perseverance in Self-perception and Social Perception: Biased Attributional Processes in the Debriefing Paradigm." *Journal of Personality and Social Psychology* 32:880–92.

Sanford, K. 2017. "Underdetermination of Scientific Theory." In *The Stanford Encyclopedia of Philosophy* (Winter 2017 Edition), edited by E. N. Zalta, at https://plato.stanford.edu/archives/win2017/entries/scientific-underdetermination/.

Shaffer, M. 2013. "Epistemic Paradox and the Logic of Acceptance." *Journal of Experimental and Theoretical Artificial Intelligence* 25:337–53.

———. 2017. "A Thoroughly Modern Wager." *Logos & Episteme* 8:207–31.

Stein, E. 1996. *Without Good Reason*. Oxford: Clarendon Press.

Steup, M. 1986. "The Deontic Conception of Epistemic Justification." *Philosophical Studies* 53:65–84.

———. 2000. "Doxastic Voluntarism and Epistemic Deontology." *Acta Analytica* 15:25–56.

Stich, S. 1990. *The Fragmentation of Reason*. Cambridge, Mass.: MIT Press.

Wason, P. 1968a. "Reasoning About a Rule." *Quarterly Journal of Experimental Psychology* 20:273–81.

———. 1968b. "On the Failure to Eliminate Hypotheses . . .—A Second Look." In Johnson-Laird and Wason 1977, 307–14.

———. 1968c. "Self-contradiction." In Johnson-Laird and Wason 1977, 114–28.

Wason, P., and P. Johnson-Laird. 1970. "A Conflict Between Selecting and Evaluating Information in an Inferential Task." *British Journal of Psychology* 61:509–15.

Wood, T., and E. Porter. Forthcoming. "The Elusive Backfire Effect: Mass Attitudes' Steadfast Factual Adherence." *Political Behavior.*