The Ontological Argument from Reason

Why Compatibilist Accounts of Reasoning Fail

Angus Menuge

Abstract:


4. Of course, the two versions of the argument from reason are closely related, and it is appeal to the inadequate ontology of naturalism that grounds epistemological versions of the argument. Interestingly enough, at the November 2010 meeting of the Evangelical Philosophical Society in Atlanta, Georgia, Alvin Plantinga presented a paper entitled “A New Argument against Materialism,” in which he argued that materialism does not have the ontological resources to explain why the content of people’s beliefs, desires and intentions is so reliably congruent with their actions (e.g., why it is that people regularly check the fridge for a beer when they want a beer and believe that there is one in the fridge). Since these mental contents are the agent’s reasons for action, this qualifies as an ontological argument from reason against naturalism.
as a result of brute, blind, natural causal forces?" Naturalistic philosophers like Searle and Nagel are well aware of this challenge, and the ontological argument from reason has received recent attention by J. P. Moreland.

Naturalism is here understood as the claim that foundationally, there are only physical objects (particles, for example) with physical properties and relations, governed by undirected forces. What makes these objects, properties and relations “physical” is that they can be fully understood by paradigmatic physical theories (such as the atomic theory of matter and Darwinian evolution), or by future (or ideal) extensions, refinements and successors of these theories in the same vein. The qualifier “in the same vein” means that whatever else is changed in physical theory, certain kinds of alleged entities and properties will remain excluded, notably, mental substances (such as Cartesian egos or a transcendent God), agent causation, and irreducible teleology, intentionality, and subjectivity. In other words, naturalism asserts that all of reality can (in principle) be understood from an impersonal perspective appealing only to aggregates of physical particles and undirected, event causal processes. This implies that anything that appears to transcend these ontological resources (such as consciousness) in fact reduces to, supervenes on, or emerges from those resources, or else is nonexistent.

The ontological argument from reason presented here centers on the relationship between human reasoning and free will. It claims to show that human reasoning requires libertarian free will, which presupposes ontological resources, such as substantial selves with active power, that cannot be located in a naturalistic world. For the naturalist, the most plausible way out of this argument is some version of compatibilism, according to which, even granted naturalistic determinism, reasoning can occur if an agent is responsive to rational causes. However, I will show that while naturalism is compatible with a passive notion of rationality (exemplified by computers), it cannot capture the active rationality of deliberation. A less popular way out is to argue that libertarian free will can be reconciled with naturalism by claiming that it is an emergent phenomenon. In this paper, I will focus on

7. It is well-known that rigorous attempts to define naturalism along these lines tend to make naturalism either obviously false or vacuous. For discussion, see my “Is Downward Causation Possible? How the Mind Can Make a Physical Difference,” *Philosophia Christi* 11 (2009): 93–110. For the purposes of this paper, we will ignore that objection and assume that a plausible, informative definition of naturalism can be stated.
8. E.g., John Searle defends this position in his *Freedom and Neurobiology* (New York: Columbia University Press, 2007). Similar views are developed by Laura Ekstrom, Robert Kane, and Timothy O’Connor; e.g., see their essays in *The Oxford Handbook of Free Will*, ed. Robert Kane (New York: Oxford University Press, 2002).
the compatibilist response, leaving the attempt to naturalize libertarian free will for another day. I will take the work of Dennett and Fischer and Ravizza as representative of compatibilism and show that their accounts of “reasons-responsiveness” cannot provide a naturalistic account of human reasoning.

**The Ontological Argument from Reason**

Consider any paradigm case of logical reasoning apparently requiring conscious deliberation. For example, suppose that Paul applies *modus ponens* to two of his beliefs, whose content is of the form if *A* then *B*, and *A*. In order for Paul to be credited with reasoning to the conclusion *B*, we must suppose that Paul is an entity with the following powers. First, Paul must own both of the beliefs that form the premises of the argument: these beliefs must be united in one consciousness. The beliefs are inseparable from Paul in the sense that they cannot exist except as modes of Paul’s consciousness: the identity of these beliefs is intrinsically tied to Paul. Of course, it may be that Jennifer has beliefs with exactly the same content, but they are still not Paul’s beliefs, and even if Paul had convinced Jennifer of the truth of his beliefs, he cannot transfer his beliefs to Jennifer. However, for naturalism, an individual’s brain is an aggregate of separable parts: these parts can exist without the whole, and can exist inside another whole as well. So, it is possible for a part of Paul’s brain to exist outside his brain, and even, theoretically for it to exist in Jennifer’s brain. Due to this basic contrast between physical aggregates and mental states, naturalism has a very hard time explaining why an agent’s beliefs (or other reasons) are inseparable from him. This is sometimes called the binding problem: if physical aggregates consist of separable parts, how can their mere recombination in different external relationships explain the occurrence of mental states which are intrinsically tied to the mental life of one subject?

Further, naturalistic science reveals a brain that processes widely distributed informational signals in parallel. There is no one entity that exists over and above the parallel streams of information processing in the brain. For example, evolutionary psychologist Steven Pinker asserts:

> “[T]here’s considerable evidence that the unified self is a fiction—that the mind is a congeries of parts acting asynchronously, and that it is only an illusion that there’s a president in the Oval Office of the brain who oversees the activity of everything.”

---


10. For a discussion of the metaphysical distinction between separable and inseparable parts, see Moreland, *The Recalcitrant Imago Dei*, chap. 5.

Likewise, Daniel Dennett compares cognition to multiple, partial narratives assembled by different sources at different times, like different trains carrying different informational freight on different railway lines. He insists that “there is no one place in the brain through which all these causal trains must pass in order to deposit their content.” But then, on a naturalistic account, how is it possible for Paul to compare the content of his beliefs and to consider their joint implications? For on this account, our various beliefs are not united in one place, so even if a consciousness is associated with each belief, there would not be one consciousness containing all the beliefs. In our example, it would be as if Jack believed that if A then B, and Jill believed that A. In these beliefs, neither Jack nor Jill has a reason to conclude that B. Plainly, there must be a single entity that unites both of the reasons at the same time. Merely to assert that such a psychological unity emerges from neurobiological disunity is implausible, nonexplanatory, and question-begging against rival dualist accounts.

Second, Paul must be able to evaluate these reasons to see what if anything follows from them. This requires a subject, a point of view on the world, capable of interpreting its beliefs and their logical implications. However, for naturalism, the brain can be completely described in impersonal terms: there is no subject, and so while there may be an impersonal transition to a state that contains a possible conclusion, there is no entity capable of seeing that state as the conclusion. We would be no different from a computer whose algorithm drives it to a terminal state which is in accordance with reasoning, but which cannot be credited with reasoning to that conclusion.

Third, in order for Paul to reason, the very same entity that owns and interprets the two beliefs must persist over the time it takes to draw the conclusion. Otherwise, it would be as if Jack believed both that if A then B and that A, but Jill believed that B. Neither Jack nor Jill could be credited with reasoning to the conclusion that B. Jack does not survive long enough to draw the conclusion, and while the conclusion occurs to Jill, she does not derive it from the premises, which she never knew. Unfortunately, this is just what naturalism predicts. For on the naturalistic understanding, there is no plausible ontological ground for a continuant self. This is because over time, the brain is in a state of constant flux: there is no “base line” brain process which persists throughout our conscious life. As neuroscientist Mario Beauregard points out,

No single brain area is active when we are conscious and idle when we are not. Nor does a specific level of activity in neurons signify

---


13. Moreland develops this argument in his critique of Timothy O’Connor in *The Recalcitrant Imago Dei*, 141.
that we are conscious. Nor is there a chemistry in neurons that always indicates consciousness.\(^{14}\)

Since the naturalistic view denies that there is a continuant self, all it can claim is that a bundle of brain states that includes one set of beliefs (the premises) is replaced by a bundle with another belief (a conclusion), but neither bundle can be said to have reasoned to a conclusion. Reasoning requires a unified, persistent self, but naturalism appears to offer only cross-sections of causal processes in which any distinction between self and not-self is arbitrary.

Not only must this self be something over and above the bundle of its reasons, it must also have libertarian free will. For if determinism is true, a decision is the passive and automatic consequence of the bundle of reasons prior to that decision. However, as Searle points out, in clear cases of deliberative reasoning, an agent’s reasons cannot be viewed as sufficient event causes of his decisions, for then there is no distinction between compulsive and noncompulsive decisions.\(^{15}\) For example, we judge that if I am so seized by a desire, for example, for double-chocolate cake, that I devour the cake like a machine, then although there was a reason for my action (my desire), my action is not the result of reasoning. Likewise, when we deliberate about something (for example, which book to read next), if our decisions arise automatically from the reasons (for and against various volumes) we had prior to deliberation, then deliberation is a pantomime that makes no difference to those decisions. Deliberation involves attending to evidence, goals and means, evaluating all these, and drawing a practical or theoretical conclusion. But as Searle argues, this process has a point only if my beliefs and desires are not by themselves sufficient to yield the decision: there is a gap between my reasons and that decision (and other gaps as well\(^{16}\)). The evidence of this gap is that we can have many reasons to act, but act on only some, or none, of them.\(^{17}\) Since there is a gap between these beliefs and desires and the decision, this gap must be bridged by something else, and the clear evidence

---

16. Searle also discusses the gap between the decision and the action (as when a student decides to get up for an 8:00 a.m. statistics class but lacks the willpower to follow through), and the gap between the initiation and continuation of an extended action (as when one tires of removing toilet paper from trees decorated by teenagers).
17. One of the most persuasive arguments for the gap is the fact that we sometimes consider various options (e.g., for an outing), and veto all of them. Only then may we consider an alternative course of action, which we end up doing. If determinism were true, we would expect that the most powerful of our reasons prior to deliberation would be victorious, compelling that decision, but this is false to experience, not only because deliberation can alter the initial weight we put on our reasons, but because, as in this example, we can reject all our initial reasons and do something else instead. The veto and selection powers of the will are good evidence that it adds something that our reasons would not do by themselves.
of introspection is that this entity is a unified, enduring self, that owns these beliefs and desires, and which uses its active power to evaluate and select some of them in making that decision. In order for the self to do this—to “act in the gap”—it must have libertarian free will, for it is the self’s free act of endorsing some reasons rather than others which explains its decisions.

So, as developed here, the ontological argument from reason against naturalism claims that reasoning requires unified, continuant, substantial selves characterized by active power and libertarian free will, and that these resources cannot be located in a naturalistic ontology. Against this argument, the naturalist compatibilist hopes to show that reasoning survives in an impersonal world of passive, automatic event causation. I will now unpack the compatibilist theory and argue that it cannot account for any variety of reasoning worth wanting.

**Compatibilism and Human Rationality**

The central compatibilist intuition is that the decision of a subject $S$ is free if it is caused in the right way, by rational causes. So if $S$ lacks free will, it is not because $S$’s decisions have prior causes outside $S$’s control, but because these causes bypass $S$’s rationality. Typical examples are drug addiction, compulsive disorders, brainwashing, and the manipulation of mad scientists. Thus the drug addict may believe that heroin is killing her, desire to live, yet, owing to a physiological addiction, still decide to shoot up, and the compulsive may know the oven is off, but check it again anyway, and so forth. For Dennett, what makes such people unfree is that their actions result from a closed program that is insensitive to relevant information. This is analogous to the Sphex wasp that can be trapped in an indefinite loop, repeatedly moving a paralyzed cricket to the threshold of its burrow and going inside to check if it is safe. All the investigator has to do is reposition the cricket a few inches from the threshold when the wasp is inside the burrow: Sphex never has the wit to drag the cricket inside. $^{18}$

By contrast, Dennett argues, free will “involves responsiveness to reasons.”$^{19}$ So, if the addict were free, her decision would respond to her reasons to quit, and if the compulsive were free, he would not bother checking the oven again because he already knows it is off. Against the libertarian intuition that our decisions are only free if they are not determined by factors outside our control, Dennett says we can still be free so long as our beliefs and desires are controlled by the right causes, and that not all cases of manipulation remove free will. While a demonic neurologist who induces irrational beliefs and fears might rob us of free will, we would remain free if

---


$^{19}$ Ibid., 28.
we were overwhelmed by the persuasive power of a well-informed, truthful oracle that cites good evidence and credible sources for making a decision.\textsuperscript{20} This holds even if there were no possible alternative choices: for Dennett, being compelled by reason is compatible with free will, so that Luther’s stand on the scriptures was free even if he really could do no other.\textsuperscript{21} And, he claims, what matters is not alternative counterfactual possibilities, but the actual sequence: “the actual causal chain of deliberation and choice running through the agent.”\textsuperscript{22}

A major worry about Dennett’s account is that he fails to explain why some particular reasoning is owned by a particular agent. A mathematician’s notepad is responsive to the rationality of the formulae he inscribes on it, but no one thinks these formulae reflect the reasoning of the notepad: the reasoning is owned by the mathematician. But then what happens if we change Dennett’s demonic neurologist example, so that a kinder, gentler neurologist implants excellent reasoning of his own in a patient, to which the patient responds by making a decision, say, to eat more broccoli. It seems on Dennett’s account that there is no principled way to distinguish this from his persuasive oracle example: we ought to say that the choice is free, since it is reasons-responsive. But our intuition is that it was not the patient’s free choice, because the choice was not based on the patient’s own reasons, but on the neurologist’s. These and other “manipulation arguments” appear to show that a person’s using reasons is not sufficient for claiming that a person is reasoning for himself. After all, a computer can use the rational algorithms of a computer program without doing any reasoning of its own.

Partly to address such worries, John Martin Fischer and Mark Ravizza develop an account sympathetic to Dennett’s intuitions, but better equipped to handle manipulation arguments.\textsuperscript{23} Fischer and Ravizza do not explicitly advance their view as a naturalistic one, claiming that they aim for an ontologically neutral account of reasons for action.\textsuperscript{24} But for our purposes, the important question is whether their account will help the naturalist compatibilist to avoid manipulation arguments. Fischer and Ravizza agree with Dennett that responsibility\textsuperscript{25} does not require control over alternative possibilities (which they call “regulative control”) but only that the actual sequence resulting in an agent’s action results from his own, reasons-responsive mecha-

\textsuperscript{20} Ibid., 65.
\textsuperscript{21} Ibid., 133. Dennett’s example is very implausible as many have pointed out. Given his awareness of credible threats of imprisonment and death, Luther had plenty of prudential reasons not to take a stand on scripture’s teaching and plausibly could have decided to wimp out. Besides that, Luther was reporting the result of making up his mind, not the process of doing so.
\textsuperscript{22} Ibid., 132.
\textsuperscript{24} Ibid., 68n11.
\textsuperscript{25} While some would argue that one can be responsible without free will, Fischer and Ravizza assert at least that they are exploring a “freedom relevant” notion of responsibility.
nism (which they call “guidance control”). They offer a more detailed account of what it means for an agent to be reasons-responsive, rejecting both strong and weak versions in favor of “moderate reasons-responsiveness.”

This is cashed out in terms of a subject’s receptivity and reactivity to reasons in actual and hypothetical cases. Receptivity to reason concerns an agent’s ability to recognize reasons, including reasons to do otherwise than they actually chose to do. Reactivity to reasons concerns the tendency to make decisions that reflect one’s best reasons in actual and possible cases. Subjects are moderately reasons-responsive when they exhibit both (a) a regular pattern of reasons-receptivity and (b) at least weak reasons-reactivity. The idea behind (a) is that we can confidently count someone as responsible only if the reasons they recognize for doing (or refraining from doing) an action form “an understandable pattern, minimally grounded in reality.” Thus, there is probably something a bit odd about a man who would sell his Packers–Bears tickets for a thousand dollars, but not for any larger amount (although one can imagine special circumstances that would explain it). Likewise, even if a person’s decisions exhibit a coherent pattern, if they are oriented around the favor and disfavor of the magic pumpkin, they may be too divorced from reality to ground a claim of responsibility. The idea behind (b) is that weak-willed or whimsical people are still responsible for choices that ignore their unopposed best reasons so long as there are some relevant possible worlds in which their decisions affirm those reasons.

Fischer and Ravizza also directly address manipulation arguments, by insisting that in order for an agent to be responsible for his choices, he must take responsibility for the mechanism that produces them. He takes such responsibility by: (1) seeing himself as the source of the decision; (2) accepting himself as a fair target for what P. F. Strawson called the “reactive attitudes” of praise and blame; and (3) basing (1) and (2) appropriately on evidence. Intuitively, the idea is that the person who (1) sees himself as choosing to accept the oracle’s deliverances, (2) accepts the praise or blame this may engender, and (3) does (1) and (2) based on appropriate evidence (for example, evidence that he has not been coerced, that the oracles is well-informed, and so on), thereby takes responsibility for basing his decision on the oracle, so that the oracle’s reasons become his reasons. But, at least assuming that the demonic neurologist was not working with the consent and encouragement of the patient, the patient never took responsibility for the neurologist’s neural intervention as a source of his own reasoning: he does not see himself as the source of the resulting decisions or an apt target

26. Ibid., 28–41.
27. Ibid., chap. 3.
28. Ibid., 81.
30. Fischer and Ravizza, Responsibility and Control, chap. 8.
of praise and blame for those decisions. So we can claim that the person following the oracle was reasoning for himself, and thus responsible for his decisions, but the patient of the demonic neurologist was not reasoning for himself and is not responsible.

Fischer and Ravizza are aware that basing the ownership of decisions on taking responsibility commits them to a kind of subjectivism defended by Galen Strawson, according to which a person who does not believe he is responsible really is not responsible: just as the captain who falsely believes his ship’s rudder is broken is not responsible for the ensuing shipwreck, the person who believes he cannot make rational decisions is not responsible for his actions even if his belief is mistaken.\(^3\) So on this subjectivist view, a slow-witted person who did not notice the persuasive power of the truthful oracle over him, but still “went with the flow,” might not be responsible for his decisions because he never recognized it as his own decision: he was a mere conduit of the oracle’s rhetoric. And on the other hand, a crafty but weak logic student who volunteered for a kinder and gentler neurologist’s rationality implants might be responsible for its results (for example, improved test scores) because he has endorsed the deliverances of the implant as a source of his own decisions.

**Critique of the Compatibilist Account of Human Rationality**

For several reasons, the compatibilist accounts of Dennett and Fischer and Ravizza fail to locate reasoning in the world described by naturalism. Sometimes, the accounts are simply inadequate, and cannot bridge the gap between naturalistic resources and human rationality. On the other hand, when the accounts do seem to work, a metaphysical “audit” reveals that they are heavily indebted to nonnaturalistic ontological commitments.\(^3\)

**Reasons-responsiveness**

The notion of “reasons-responsiveness” is ambiguous between an active notion, which is relevant to reasoning but unavailable to the naturalist, and a passive notion, which is available to the naturalist but is insufficient for reasoning. On the passive view, any system which is governed by reason, and capable of producing outputs in accordance with reason, is reasons-responsive. This notion is certainly compatible with naturalistic determinism.


\(^3\) To be fair, since Fischer and Ravizza do not (like Dennett) explicitly tie their account to naturalism, one might conclude that to the extent that their account is both plausible and requires nonnaturalistic ontological commitments, it can be used as part of an extended argument against naturalism.
because a natural system might be so configured that its causal processes mirror reason. In this sense, any pocket calculator is reasons-responsive. Yet no one thinks that the calculator is reasoning for itself, and Fischer and Ravizza would certainly deny that it is, because the calculator is not capable of taking responsibility for its “decisions.” However, the active notion, according to which the system does reason for itself, is unavailable to the naturalist. For according to naturalism, there are no mental substances with active power: all causation is automatic, passive event causation. So there is no mental entity over and above the various physical, causal processes in the brain that can take responsibility for some of those processes as its own reasoning. More fundamentally, before an entity can be responsible for its reasoning, we need an account of what makes that reasoning belong to it, rather than merely being some reasoning that is going on in a certain region of space and time. Yet, as already noted, on a naturalistic view, there is no entity that can plausibly own any mental states: there is simply a plurality of parallel processes in the brain. This problem can be sharpened with a couple of examples.

Case 1. Suppose that Albert believes that \( A = B \) and also that \( B = C \), and as a result of these two beliefs concludes that \( A = C \). Albert’s reasoning makes sense if there is some one mental substance (Albert) that owns and unifies the beliefs that \( A = B \) and \( B = C \) in one consciousness, and that endures over the time it takes to draw the conclusion that \( A = C \). But if there are no mental substances, then even if one brain process contains the information that \( A = B \) and another contains the information that \( B = C \), there is no entity that unites the information in one act of thought at a time, or which can persist to draw a logical conclusion over time. Should the information that \( A = C \) arise in some later brain process, there is nothing that could be credited with having reasoned to that process. Without a unified, enduring subject, compatibilist naturalists have no way to distinguish Albert’s act of reasoning from many other causal processes that would produce a rational conclusion without reasoning, such as those at work in digital computers.

Case 2. Suppose that Jeff believes that \( A = B \) and that \( B = C \), but is so lethargic that he does not draw the obvious conclusion. Happily, a kind neurologist monitoring Jeff’s brain sees the information there and induces him to believe that \( A = C \). Then Jeff has reached a conclusion in accordance with reason, but he has not done so by reasoning. Fischer and Ravizza can respond that Jeff did not take responsibility for the process that produced his belief, but this does not help the naturalist because naturalism does not explain the existence of any entity that could take such responsibility. For if taking responsibility simply reduces to (or supervenes on, or emerges from) another brain process, then it must be possible (at least in theory) for the neurologist to trigger that process as well, making it impossible to distinguish genuine cases of taking responsibility from brainwashing. To make this distinction,
therefore, one must distinguish between cases where Jeff is the patient of someone else’s actions and cases where Jeff exercises his own active power. But naturalism does not permit active power: it is a characteristic of mental substances with libertarian free will.

**Unstoppable Manipulation and the Ratiomaniac**

This last point can be developed to show that even with its many ingenious conditions designed to block cases of manipulation, Fischer and Ravizza cannot avoid all manipulation arguments. Suppose Jill, a logic professor and former neuroscientist, is exasperated with her weak student Jack. So Jill implants a rationality enhancer in Jack’s brain that not only makes Jack moderately reasons-responsive but also induces him to take responsibility for the decisions based on the rationality enhancer, by making him endorse the enhancer as part of himself, so that he sees himself as the source of the decisions it produces and an apt target for any praise or blame this provokes, all based on the apparent evidence that the enhancer is fully integrated with himself. In this case, Jack satisfies all Fischer and Ravizza’s conditions for responsibility. But arguably, although he thinks he is reasoning for himself, he is not. Should the implant contain a bug, Fischer and Ravizza would have to say incorrectly that Jack was responsible for the error of reasoning. But not only is Jack not responsible in this case (perhaps Jill is), he is not reasoning at all: he is simply the host of a rational parasite that has successfully coerced his endorsement of the parasite’s “decisions” as his own, rather like individuals who have been “converted” by the cybermen on the science fiction show *Doctor Who*. These individuals continue to believe that they are reasoning for themselves, but all their decisions result from an implanted program that automatically implements the cybermen’s agenda. The general problem is that, if determinism is true, it must be possible to make someone take responsibility for an alien mechanism (with alien reasons) that causes their decisions.

It may be objected that the compatibilist should simply bite the bullet, and claim that, in our scenario, Jack is reasoning for himself and is responsible for his decisions. Indeed, this is the line that Fischer takes in response to some of Derek Pereboom’s examples of manipulation.33 However, in our present case, the price of biting the bullet seems unacceptably high, because it means that the compatibilist is no longer able to draw the principled distinction between those decisions that are compulsive or coerced, and those that are not, which motivated the account in the first place. The compatibilist began with examples like drug addiction, where all parties agreed that there

---

was a deficiency in responsibility. But the present account implies that if one is addicted to moderate reasons-responsiveness and to taking responsibility for the decisions resulting from one’s addiction, then one must be counted fully responsible after all. It appears that biting the bullet in one case will lead to a steady diet of lead poisoning.

But perhaps interventions like implants and drugs seem artificial and could be blocked by the compatibilist in some other way. No matter, the same problem can be raised without such props. It is perfectly conceivable that a human being suffers from compulsive moderate reasons-responsivity and also compulsively takes responsibility for his decisions. In his classic work, *Orthodoxy*, G. K. Chesterton unkindly compares the modern materialist to a maniac who has lost everything except his reason and so cannot stop thinking around the same closed circle. In the ratiomaniac’s mind, one might say that reasoning occurs and carries on automatically, yet it seems wrong to say that the ratiomaniac is reasoning because the ratiomaniac has no power to stop thinking along this trajectory. A good example might be R. M. Hare’s famous example of the “paranoid man,” who thinks obsessively about the possible dangers of every apparently kind action: “Mightn’t the cupcakes be poisoned?” “You are just being nice to lull me into a false sense of security.” And so on. Does that mean the paranoid ratiomaniac is not moderately reasons-responsive? Not necessarily. It is surely possible that the ratiomaniac’s recognition of reasons follows a coherent, understandable pattern (regular reasons-receptivity) and that sometimes, his decisions would affirm those reasons (weak reasons-reactivity). Isn’t it quite predictable that he will think of possible hidden dangers and threats, and make precautionary decisions on that basis? And isn’t it also possible that the ratiomaniac takes responsibility for his decisions, and, for example, isn’t surprised when friends criticize him for being “gloomy,” or when security professionals thank him for spotting potential security holes at airports? So on Fischer and Ravizza’s account, the ratiomaniac is apparently responsible for his decisions, yet it may be that he is in the grip of a highly sophisticated neurotic program, and is no more reasoning for himself, than is the obsessive compulsive individual who finds himself checking yet again if the oven is off.

Four objections to this example are likely from the defender of Fischer and Ravizza’s account. First, it might be argued that there is something not fully intelligible about the ratiomaniac’s decisions. But this seems unlikely, because the ratiomaniac’s decision-making mimics that of psychologically normal trained security professionals who are reasoning for themselves: it is just that, unlike him, they can stop thinking that way when they are not on

---

the job. Second, it might be claimed that the ratiomaniac is not moderately reasons-responsive because he does not recognize reasons to do otherwise than make the paranoid decisions that he does. In fact, this need not be true. Consider the case where the ratiomaniac decides not to eat the proffered cupcakes for fear that they are laced with poison. It is certainly possible that he would recognize some reasons to eat the cupcakes. All he would need is a stronger, competing source of paranoia. For example, suppose that when the ratiomaniac says, “No, thank you,” the host grimaces, scans the antique military saber mounted on his wall, and makes a joke about using it on rude guests. The ratiomaniac may now think that the consequences of not eating the cupcakes could be serious injury or death, and so may decide to risk the possible poison. The availability of this sort of response is guaranteed by the fact that paranoia is a global perspective (a “blik,” in Hare’s parlance) that colors all possible evidence: all that is needed to overturn a paranoid reason not to do $A$ is a stronger paranoid reason to do $A$. Thirdly, it might be claimed that the ratiomaniac is “substantially deluded” about the nature of reality, and so the compatibilist can claim that he is not responsible. But in fact, it is not clear that he is substantially deluded, or that if he is, that he is so in a way that allows one to claim that the pattern of his reasons is irrational. The possible dangers or threats on which the ratiomaniac fixates need not be logically absurd or excluded by known fact. Indeed, he need not even seriously believe that cupcakes are poisoned, and so on; it is just that he thinks of and worries about that sort of possibility when others do not. Further, if he is substantially deluded, it seems one would have to claim erroneously that so is the trained security professional who is paid to reason out ways to defend against just such possible but unlikely threats. More generally, if Chesterton and Hare are correct, materialists themselves are much like the ratiomaniac, yet, supposing that they are deluded, it is surely not in such a way as to preclude their reasoning for themselves or being responsible. So monster-barring attempts to exclude the ratiomaniac would likely prove too much, excluding clear cases of reasoning and responsibility. The last response is yet again to bite the bullet, but this means that the principled distinction between compulsive and noncompulsive decisions that was used to motivate the compatibilist account is no longer available.

My conclusion from all this is that it is just a matter of ingenuity to describe a sophisticated compulsive reasoner who will fulfill all Fischer and Ravizza’s conditions without reasoning for himself. This is just what we would expect given compatibilist naturalism, because it only recognizes the automatic, passive event causation of efficient causality: it must be possible for its theory of rationality to be automated in a system that is not reasoning

36. Fischer and Ravizza insist that their account presumes some external constraints on rationality, so that “the agent not be substantially deluded about the nature of reality” (Responsibility and Control, 73).
for itself. So the apparent problem for compatibilist naturalism is that all reasoning reduces to compulsive reasoning, and no one can transcend the limitations of the ratiomaniac. This provides a good motivation to invoke a mental entity with libertarian free will, which allows us to make a clear distinction between compulsive and noncompulsive reasoning.

**Guidance Control**

It might seem that Dennett and Fischer and Ravizza have a ready response to this sort of objection because their central idea is that responsibility implies control. Surely, one might argue, the ratiomaniac is not in control of his reasoning but is like a driver who cannot stop his car because the accelerator is jammed. However, it can be shown that the kind of control available to the naturalist is not sufficient to exclude the ratiomaniac, and the kind that is sufficient is not compatible with naturalism. For Fischer and Ravizza, “control” means guidance control.\(^{37}\) According to this idea, if \(S\) is driving a car and decides to turn left, \(S\) can be said to freely guide the car to the left even if \(S\) does not have the power to guide the car to the right. That is, \(S\) can have guidance control without the ability to regulate between alternative possibilities (regulative control). The idea is that the famous consequence argument shows that \(S\) did not have the power to do otherwise (turn right), but it is compatible with this that \(S\) can guide the actual sequence (turn left), and this power is enough for responsibility (and, arguably, even some notion of free will).\(^{38}\)

But what does guidance mean? Intuitively, a rational person guides his decision-making by reflection on his own goals: if such reflection makes no difference to what a person does, then he is like the ratiomaniac and we do not credit him with reasoning, despite the reasoning going on in him. However, if reflection on goals affects someone’s decisions, then goal-directed causal processes must be admitted, yet naturalism excludes teleological causation on principle. This is not surprising, because, without a mental substance, there is no entity capable of guiding itself by its own goals. So, for the naturalist, “guidance” will have to be reinterpreted in nonteleological terms.

Thus for Dennett, a system which acts outside the direct control of others and whose behavior is modified by its own feedback loop has “self-control.”\(^{39}\) But this is not enough for what we normally mean by responsibil-

\(^{37}\) Ibid., 30–41.

\(^{38}\) The consequence argument was first given a crisp statement by Peter van Inwagen in his *An Essay on Free Will* (Oxford: Clarendon, 1983), chap. 3. Put briefly the argument asserts: (1) If determinism is true, then our actions are the result of events in the remote past and the laws of nature; (2) The remote past and the laws of nature are not up to us; so (3) Our actions are not up to us: we have no power to do otherwise.

\(^{39}\) See Dennett, *Elbow Room*, chap. 3.
ity. All of it can be realized in a guided missile that changes its course depending on real-time signals from its environment. The kind of control that the missile has of its own course does not convince anyone that the missile is responsible for the reasoning resulting in its targeting “decisions,” while we do hold the designers and launchers of the missile responsible, precisely because (outside philosophy) we think that the behavior of this missile traces, at least in part, to their free, goal-directed reasoning, even if they are not able to predict the precise course of the missile. In Searle’s terminology, the missile is only rational in a derived, “observer-relative” sense (that is, relative to human goals, there is a reason the missile behaves as it does), which is his judgment about all current computers. This sort of nonteleological guidance control is not enough to show that a system is reasoning for itself.

**Ultimate Responsibility**

Naturalism is also in conflict with Fischer and Ravizza’s claim that a self becomes responsible by taking responsibility, by coming to see itself as the source of its behavior. Either the self is the ultimate source of its behavior or it is not. But granted naturalism, it is not. For if the self were the ultimate source of its behavior, it must be able to initiate causal chains, which requires a mental substance with active power. But on a naturalistic account, what we call the self is merely a cross-section from an extended bundle of passive event causal processes leading from the nonself through that cross-section to certain behavior. The only kind of responsibility this will justify is causal responsibility: the “self” is responsible for its behavior in the way a boulder is responsible for damaging a house, though the boulder did what it did as the passive, automatic result of the prior earthquake, and the “self” did what it did as the passive, automatic result of the nonself processes. Fischer and Ravizza realize that such causal responsibility is not sufficient for personal (intellectual or moral) responsibility, yet it is the only kind that seems compatible with naturalism.

**A Non-Humean Self with Libertarian Free Will**

As we saw, some naturalists, like John Searle, concede this point and argue that in order to explain rational deliberation, we must postulate a non-Humean, enduring self. Dennett tries to avoid this conclusion by arguing that assuming determinism, we can justify deliberation in the same way that the golfer justifies keeping his head down when putting, even though it can have no influence on the ball after it leaves the club: “looking ahead in this way

and resisting distractions should tend to ensure that the right, desirable sorts of things happen at the crucial time.” However, either this “looking ahead” and “resisting distractions” makes a difference to what an agent’s reasons do or it does not. If it does not, it is a superstitious pantomime, like a rain dance. If it does, as Dennett clearly seems to believe, then the self does add something to what those reasons would do by themselves, and so, as Searle argues, there is a gap in which the self acts; but this requires that self to possess libertarian free will. Searle concludes that the only hope of reconciling naturalism with human reasoning is to show that non-Humean selves with libertarian free will can emerge from the brain, which is a very tall order.

**Conclusion**

The ontological argument from reason purports to show that deliberative reasoning cannot occur in a world exclusively populated by a naturalistic ontology because such reasoning requires a unified, enduring self with libertarian free will. To avoid this argument, most naturalists will follow Dennett and Fischer and Ravizza, and attempt to show that some version of naturalistic compatibilism suffices to account for reasoning, because even in a world of event causation, some creatures may be responsive to reason. But, amongst many other problems, such an account fails to distinguish compulsive rationality that is merely occurring in someone’s brain from reasoning that an agent does. Given the strength of Searle’s arguments for an irreducible self with libertarian free will, some form of dualism seems unavoidable to account for reasoning, and, though I have not attempted to exclude several rival dualist accounts, I have argued that substance dualism is up to the job.

---

43. For a searching critique of Searle’s position, see Moreland, *The Recalcitrant Imago Dei*, chap. 3.