

Don't Think Twice, It's Alright

Hilary Kornblith

University of Massachusetts, Amherst

We arrive at most of our beliefs unreflectively. As we navigate the world, beliefs about our surroundings are, inevitably, simply produced in us. Similarly, the vast majority of our actions are unreflective. We don't have to think about every little thing we do; we simply act. But we also, at times, stop to reflect: Is this what I should believe? Is this what I should do?

What does such reflective activity achieve? Some philosophers have suggested that reflecting about what we should believe is necessary if our beliefs are to be justified. In the case of action, some philosophers have suggested that reflecting about what one should do is necessary for freedom of the will and moral responsibility. One might think that there are more humble benefits as well. Beliefs which are the product of reflective activity are more likely to be true than beliefs unreflectively arrived at; actions reflectively produced are more likely to be successful in achieving their goals than unreflective actions. This is just, it seems, good common sense.

This paper challenges both common sense views about the benefits of reflection as well as a good deal of recent philosophical thinking. It would be silly to think that reflection is never valuable, but I will argue that both common sense, and much philosophical thought about the nature and importance of reflection, have vastly overestimated its value. I will begin with an account of what I take to be the common sense view about the value of reflecting on our beliefs and actions, as well as traditional philosophical thought on this topic. My own views on this issue have become increasingly radical over recent years. I want to present an argument, step by step, both to explain, and to justify, my ever increasing radicalization on this topic. As I now see it, the common sense view not only grossly misrepresents the value of reflective belief and action; it presupposes a view about the nature of our thought which is almost certainly false. If I am right, the very nature of human thinking is nothing like the way we represent it to ourselves when we deliberate about what to believe and what to do.

1. Common Sense about Reflection

Let me begin with some common sense thoughts about human and animal minds and the differences between them.

Human beings are not the only animals with mental states. A great many other animals must be credited with beliefs and desires, as well as a variety of other mental states, if we are to explain the sophistication of their problem-solving behavior. But while dogs and cats, cows and kangaroos, parrots and chimpanzees, all clearly have beliefs and desires, there is one very important respect in which their beliefs and desires differ from ours: their beliefs and desires are directed exclusively at the world around them—they have beliefs about food and shelter, other creatures, and the various threats and opportunities which their environment presents them with—while our own mental states are not only directed at the world around us; they are also turned in upon themselves. We not only have beliefs about the world around us; we have beliefs about our mental states. We are thus capable of thinking about our beliefs and desires, and, inevitably, when we do this, we sometimes come to wonder: Should I really believe this? Should I really act on this desire? Other animals, who lack the capacity to reflect on their own mental states, do not and cannot entertain such questions.¹

We not only differ from other animals in this respect. We differ from very young children in this respect as well. While physical object concepts are early arrivals in human conceptual development, mental state concepts arrive much later. Young children have beliefs and desires directed at the world around them well before they are capable of entertaining thoughts about their own mental life or the mental lives of those around them.²

¹ I have argued that this view of the difference between human adults, on the one hand, and other animals and young children, on the other, is roughly correct in *Knowledge and its Place in Nature*, Oxford University Press, 2002, Chapter 2. For some standard works in cognitive ethology which support this view, see John Alcock, *Animal Behavior: An Evolutionary Approach*, 10th edition, Sinauer Associates, 2013; Colin Allen and Marc Beckoff, *Species of Mind: The Philosophy and Biology of Cognitive Ethology*, MIT Press, 1997; Daniel Povinelli, *Folk Physics for Apes: The Chimpanzee's Theory of How the World Works*, Oxford University Press, 2000; Jacques Vauclair, *Animal Cognition: A Introduction to Modern Comparative Psychology*, Harvard University Press, 1996.

² See, for example, Janet Wilde Astington, *The Child's Discovery of the Mind*, Harvard University Press, 1993; Janet W. Astington, Paul L. Harris, and David R. Olson, eds., *Developing Theories of Mind*, Cambridge University Press, 1988; Alison Gopnik and Andrew Meltzoff, *Words, Thoughts, and Theories*, MIT Press, 1997; and Henry Wellman, *The Child's Theory of Mind*, MIT Press, 1992.

Young children and non-human animals thus take on a rich network of beliefs about the world around them without being capable of reflecting on the questions of whether they should believe as they do, whether they should want the things they want, or whether they should act on the basis of particular desires they happen to have. Beliefs and desires are produced within them by mental mechanisms of which they are entirely ignorant. In the case of their beliefs, some of these mental mechanisms are extremely reliable: they tend to produce accurate beliefs about the world, beliefs which are far more often true than false. But some of these mental mechanisms are not reliable at all: they tend to produce far more false beliefs than true ones. These creatures, however, do not have the conceptual capacity to think about these matters. Their belief-producing mechanisms hum along, doing what they do, producing a variety of beliefs, both true and false. It is much the same with desires. A variety of mental mechanisms produce desires in children and other animals, and these desires interact with their beliefs to produce behavior. Some of these behaviors are extremely effective in allowing these creatures to get what they want, all things considered. Others are not nearly so effective. But these mechanisms, like the mechanisms which produce beliefs, simply hum along, doing what they do, producing desires and behaviors of all sorts.

Adult human beings are more complicated creatures. Our capacity to reflect on our beliefs and desires affords us a degree of sophistication that other creatures lack. It allows us to contemplate epistemological questions—questions, for example, about whether we should believe as we do—and various moral questions as well—questions about whether we should act on the basis of certain desires we may have, and even questions about whether we should want some of the things we want. More than this, it seems that we can do far more than simply raise these questions. Our ability to think about whether we should believe as we are inclined to do, and desire as we are inclined to do, and act as we are inclined to act, allows us to play a role in determining what we believe, and what we desire, and what we do that more conceptually limited creatures cannot.

Let us consider a few examples. Suppose I come to realize that I have certain beliefs about the upcoming presidential election. I believe that one particular candidate, widely discussed as a serious contender, is likely to be pushed out of the race fairly quickly. I believe that another candidate, who is not much discussed, is a far

more serious contender. I believe that one of the two major parties has little chance of winning the election, and that the other party is overwhelmingly likely to win.

Now very young children and other animals, of course, do not have beliefs about elections. But what I want to focus on here is not the fact that I have certain beliefs about the coming presidential election. Rather, I want to focus on the fact that I am aware that I have those beliefs: I not only believe certain things about the coming election; I also know that I have those very beliefs. Young children and other animals, as we've noted, do not have the conceptual capacity to form beliefs about their own beliefs. So even in the case of very simple beliefs—say, beliefs about objects in front of them—young children and other animals do not have the conceptual capacity to recognize that they have those beliefs. And my noting that I have certain beliefs is not idle; it can prompt me to think about whether I ought to have those beliefs. On coming to realize that I've already come to believe certain things about the forthcoming election, I can't help but wonder whether I am really justified in believing as I do. Are my beliefs really supported by the evidence I have, or are they a product, perhaps, of wishful thinking? When I think about these questions, I am led to think about the evidence I have for my beliefs. I do my best to evaluate that evidence to keep myself honest. If the evidence really does support my beliefs, I should continue to hold them. If it doesn't—if, for example, I come to suspect that I have been engaging in wishful thinking—then I should revise my beliefs accordingly. Beliefs are produced within me in my unreflective moments. Stopping to reflect on these beliefs and the evidence I have for them serves as a check on the processes of belief-acquisition which operate within me prior to reflection. If I don't ever stop to reflect on these beliefs, I am, in effect, just crossing my fingers and hoping for the best: hoping, that is, that the processes of belief-acquisition which operate within me when I am unreflective are ones which tend to produce true beliefs. If I do stop to reflect on my beliefs, however, I can do better than this. I can take charge of my intellectual life rather than be a mere bystander. I can make sure that my beliefs are reliably formed rather than simply trust to luck. I can be a responsible epistemic agent rather than someone who idly sleepwalks through his intellectual life.

We not only take charge of our intellectual lives when we stop to reflect on beliefs already formed in us. We may also stop to deliberate about what to believe on some subject prior to having

any beliefs on the matter at all. A student asks me for some advice on a complicated issue, one I had not given any prior thought to at all. I stop to think about the matter, and consider various courses of action. Which of the various courses of action open to her do I believe to be the best? I think about the advantages and disadvantages of various choices she might make, and I come to an opinion on the matter. Here my deliberation concerns what to believe, rather than whether a belief I already have is one I am entitled to. But just as in the case of deliberating about the credentials of an already existing belief, my deliberation about what to believe allows me to take charge of my intellectual life. I am not merely passive in allowing beliefs to be formed within me by whatever psychological mechanisms happen to be housed within my mind. I take responsibility for my beliefs by way of my deliberation. I am an epistemic agent, rather than a leaf blowing in the wind.

The same is true of my desires. I like coffee. I like coffee a lot. At a certain point, I notice that I'm drinking an awful lot of it. I have coffee before I leave the house. I make a small pot of coffee when I arrive at the office, and I finish it off before lunch. I have coffee with lunch, and then, when I get back to the office, I make another small pot of coffee, which I finish off before I return home. My coffee consumption slowly but surely increased over time without my fully realizing it. But now that I do realize it, it occurs to me that this cannot be healthy. I like coffee. I want to drink it. My coffee drinking behavior was brought about by my desire for coffee. But in reflecting on just how unhealthy my coffee consumption has become, and how my desire for coffee has slowly but surely brought this on, I come to think that I would be better off if I didn't like coffee so much. Although I like coffee, I come to wish that I didn't like it quite so much. Although I desire coffee, I desire that I should stop desiring coffee. Although I act on the basis of my desire for coffee, I desire that my desire for coffee should not be effective in producing action.

These are the kinds of thoughts that young children and other animals simply cannot have. They do not have the conceptual capacity to think about their desires, and so they cannot have these thoughts. Their behavior is brought about, in a fairly direct way, by their beliefs and desires. But I can stop to reflect on my desires, and think about whether these are desires I want to have, and whether I want the desires I have to be effective in producing action. Instead of behaving in ways that are determined by various mental mechanisms operating in me without my knowledge, I can

take charge of my life by reflecting on my desires and the kind of life I want to live. I can play an active role in determining which of my desires I act upon, rather than passively sitting back and letting my behavior simply happen. I can take responsibility for my life. I can be a morally responsible agent.

So what does the capacity to reflect upon our mental states do for us? As common sense would have it, this capacity does not merely make us more complicated than young children and other animals. Instead, the capacity to reflect upon our beliefs and desires is transformative. Young children and other animals have a great storehouse of beliefs about the world around them, and many of these beliefs are true. More than this, it is no coincidence that so many of their beliefs are true. These beliefs are produced by reliable cognitive processes operating within them. But these are not the only kinds of cognitive processes which operate in young children and other animals. Their mental mechanisms include extremely unreliable processes of belief acquisition as well. And this, in turn, means that they also have a great many beliefs which are false, and not coincidentally so. The processes of belief acquisition which operate within these creatures, processes which are shaped by natural selection, are not all geared to producing truths. All of these processes, both the ones which are extremely reliable and the ones which are extremely unreliable, operate within these creatures without their being aware of them. These creatures have a mental life, to be sure; but in one very straightforward sense, their mental life is not their own. It goes on in them, but they do not direct it. They are merely passive vessels for a complicated set of mental mechanisms and the mental states which they produce.

The capacity to reflect on one's mental life is transformative, not because it makes us more complicated, but because it allows us to be in charge of our mental lives. By reflecting on our beliefs and desires, we may take responsibility for what we believe, and what we desire, and what we do. We can play an active role in our mental lives, instead of being simply passive. The capacity to reflect allows us to be epistemically and morally responsible agents, and not merely the theater in which our mental life plays out as we stand idly by.

This is, as I see it, the common sense view of the difference between young children and other animals, on the one hand, and adult human beings, on the other. This is common sense made more explicit than it usually is, but it is, nonetheless, I believe, the common sense view of the importance of reflection.

2. Some Recent Philosophical Thinking about Reflection and Justified Belief

A great deal of recent philosophical thinking about reflection merely expands upon these common sense thoughts.

Consider, for example, Laurence Bonjour's views about the nature of justified belief. Bonjour sees a deep connection between our capacity to reflect on our beliefs and the very possibility of having beliefs which are justified. He has offered a very widely discussed counterexample to views of justified belief which divorce our capacity for justified belief from the capacity to reflect. Thus, consider reliability views of justification. On such views, a belief is justified just in case it is reliably produced.³ One need not know that one's belief is reliably produced, or believe that it is reliably produced, or even have evidence that it is reliably produced in order for the belief to count as justified, according to such views. But Bonjour argues that any such view will have deeply counter-intuitive consequences. Thus, consider Norman, who is in fact clairvoyant, but has no evidence whatsoever that he has this power. One morning, he wakes up and finds himself believing that the President is in New York. Although he has no evidence whatsoever that this is true, his belief is, in fact, caused by his reliable clairvoyant power. Reliability accounts of justified belief would need to say, it seems, that Norman's belief is justified. But as Bonjour argues, this is extremely implausible. As Bonjour remarks,

...Norman's acceptance of the belief about the President's whereabouts is epistemically irrational and irresponsible, and thereby unjustified... Part of one's epistemic duty is to reflect critically upon one's beliefs, and such critical reflection precludes believing things to which one has, to one's knowledge, no reliable means of epistemic access.⁴

So, as Bonjour sees it, when Norman takes on the belief about the President's whereabouts uncritically—that is, without reflecting on the question of whether he has adequate evidence for the belief—he is failing in his epistemic duty; he is behaving in a way which is

³ The locus classicus here is Alvin Goldman, "What is Justified Belief?," in George Pappas, ed., *Justification and Knowledge*, Reidel, 1979, 1-23.

⁴ *The Structure of Empirical Knowledge*, Harvard University Press, 1985, 42.

epistemically irresponsible; and, for that very reason, his belief about the President is unjustified. This claim, I believe, has strong intuitive appeal, and it fits well with the common sense views of reflection just canvassed.

Notice that Bonjour's objection to counting Norman's belief as justified does not depend on the fact that, in the imagined example, Norman has no evidence that the President is in New York. Rather, the problem, as Bonjour sees it, is that Norman failed to reflect on the epistemic credentials of his belief. Even if Norman had adequate evidence that the President is in New York, his failure to reflect on whether he had good reason for his belief shows him to be epistemically irresponsible in holding it, and, for that very reason, his belief is not justified. We might summarize Bonjour's view in the form of a slogan: 'The unreflected-upon belief is not worth having. But intuitive as Bonjour's reflective requirement on justified belief may be, it cannot be right.

Thus, imagine a case in which I have wonderfully detailed evidence that the President is in New York. I know that he announced a trip to New York more than a week ago; I know that the newspapers reported this morning that he will be arriving in New York shortly; I myself am in New York and I see the Presidential motorcade proceeding down Fifth Avenue. I come to believe, as a result of all of this, that the President is now in New York, and, in attempting to honor Bonjour's requirement that, as a responsible epistemic inquirer, I reflect on the credentials of my belief, I do exactly that: I ask myself whether I have good reason for my belief about the President's location, and I self-consciously scrutinize all of my evidence, and I note that my evidence provides powerful reason to believe that the President is in New York. So it seems that I have satisfied Bonjour's reflective requirement for justified belief.

But I haven't satisfied Bonjour's requirement. I have reflected on my reasons for believing that the President is in New York, and, as a product of that reflection, I came to believe that I had adequate evidence for this belief; I based my belief about the President's whereabouts on the results of my reflective examination of my evidential situation. But now consider my belief that my evidence supports my belief that the President is in New York. I haven't reflected on *that* belief. It was formed as a product of reflection, but it has not yet been reflected upon. And Bonjour says that I have an epistemic duty to reflect on my beliefs, and, if I hold some belief that I have not reflected upon, I am being epistemically ire-

sponsible, and, so, that belief is not justified. Since I haven't yet reflected on my belief that I have sufficient evidence to believe that the President is in New York, that belief, according to Bonjour, must count as unjustified. And since my belief about the President's whereabouts is based on my belief about my evidential situation, which is, according to Bonjour, unjustified, it follows that my belief about the President's whereabouts is also unjustified.

The problem, I think, is quite straightforward. According to Bonjour, if I have not reflected upon my reasons for holding some particular belief, then I am being irresponsible in holding that belief, and so that belief counts as unjustified. So I stop to reflect. But now the beliefs formed as a product of reflection have not themselves been reflected upon, and so they are unjustified. If I stop to reflect on them, then the same problem arises all over again. Reflection must stop somewhere, and wherever it does, the beliefs which result from it are themselves unjustified. But since the point of reflection was to justify my original belief, and that belief is now based on an unjustified belief, it too is unjustified. Bonjour's reflective requirement is thus unsatisfiable. It sets a standard for justified belief which cannot possibly be met. The suggestion that a belief cannot count as justified unless it is reflected upon thus cannot be right.

One might think that there is still some reason to connect justified belief with reflection, even if the reflective requirement cannot be put in the very simple and straightforward way that Bonjour presents it. Surely, one might say, there are advantages to be gained by reflecting on one's beliefs, even if one cannot reflect "all the way down," as Bonjour's requirement would have us do. Ernest Sosa offers a very plausible suggestion. Unlike Bonjour, Sosa does not require that a belief be reflected upon in order to count as justified. He does, however, suggest that reflective knowledge is better than knowledge without reflection, that is, it is better than what he calls "animal knowledge."

One has *reflective knowledge* if one's judgment or belief manifests not only...[a] direct response to the fact known but also understanding of its place in a

wider whole that includes one's belief and knowledge of it and how these came about.⁵

If one stops to reflect on one's belief—say, that the President is in New York—one may come to be aware of a variety of facts about its evidential standing and its causal origin, things one would not have been aware of without stopping to reflect. And on Sosa's view, while these things are not necessary for justified belief or for knowledge, they serve to improve one's epistemic position.

Since a direct response supplemented by such understanding would in general have a better chance of being right, reflective knowledge is better justified than corresponding animal knowledge.⁶

So the thought here is that reflection can make one better justified than one would have otherwise been by making one aware of one's evidential situation, and, in virtue of that fact, beliefs reflected upon are more likely to be true than beliefs which are not reflected upon; they “have a better chance of being right.”⁷

The claim that reflecting on the epistemic credentials of one's beliefs improves one's chances of having true beliefs is utterly commonsensical. It is, of course, an empirical claim about the power of reflective self-examination, and so it is worth examining the empirical evidence here to see what is known about the effects of such reflection. As it turns out, however, the empirical evidence does not support Sosa's view. Let me briefly review some of the relevant evidence.⁸

When we stop to reflect on our reasons for belief, we are not capable of bringing to mind all of the evidence we have. The capacity of short-term memory is quite small, and this limits the

⁵ “Knowledge and Intellectual Virtue,” reprinted in his *Knowledge in Perspective: Selected Essays in Epistemology*, Cambridge University Press, 1991, 240.

⁶ *Ibid.*

⁷ In more recent work, Sosa has not made this claim. See, for example, his *Reflective Knowledge: Apt Belief and Reflective Knowledge*, vol. II, Oxford University Press, 2009; *Knowing Full Well*, Princeton University Press, 2011; and *Judgment and Agency*, Oxford University Press, 2015.

⁸ I have provided a more thorough review of the evidence in *Knowledge and its Place in Nature*, chapter 4, and in *On Reflection*, Oxford University Press, 2012, chapter 1.

number of items we can bring to mind at any one time. On the whole, it is often easier to remember recent experiences than ones from our distant past, and a variety of factors which are entirely unconnected to the strength of one's evidence plays an important role in determining how easy it is to bring these bits of evidence to consciousness. As a result, the evidence we bring to mind when we reflect is unlikely to give us a terribly accurate picture of the evidence we actually have which bears on our beliefs. This may make it sound as if it is largely a matter of random factors at play when we reflect whether we will come to overestimate the strength of our evidence, or underestimate the strength of our evidence, or, by mere chance, get it roughly right. But, as it turns out, things are even worse than this. There is a very well documented confirmation bias at work when we reflect. Beliefs which support the belief reflectively examined are more easily brought to mind than disconfirming evidence. Thus, when we stop to reflect on the evidential credentials of one of our beliefs, we are likely to find that the evidence we self-consciously survey gives us good reason to go on believing as we already do. Reflectively surveying our evidence does not serve to alert us to our mistakes; it does not make our resulting beliefs more accurate, on the whole, than they would otherwise be. What it does, however, is tend to increase the confidence we have in the beliefs we antecedently held.⁹

The depth of the problem here is best illustrated by the phenomenon of confabulation. Let us consider a particularly striking demonstration of this phenomenon, the case of choice blindness. Subjects are asked a series of questions about their preferences or their beliefs. They may be asked, for example, whether they prefer strawberry or grape jam; they may be asked whether they favor raising taxes or lowering them; they may be asked about their opinions on a variety of moral and political issues. Having answered these questions, the subjects are then asked why it is that they chose as they did. Subjects are divided into two groups. The first group is asked, for each choice which they in fact made, why they made that choice. But in the second group, subjects who chose, for example, strawberry over grape jam are asked, "Why did

⁹ See, for example, Peter Wason and Philip Johnson-Laird, *Psychology of Reasoning: Structure and Content*, Harvard University Press, 1972; Richard Nisbett and Lee Ross, *Human Inference: Strategies and Shortcomings of Social Judgment*, Prentice-Hall, 1980, chapter 8; Hugo Mercier and Dan Sperber, *The Enigma of Reason*, Harvard University Press, 2017

you choose grape jam rather than strawberry?’. In each case in the second group, subjects are asked, that is, why it is that they made a choice which, in fact, they didn’t make.

Amazingly, only a very small percentage of subjects notice that they are asked to explain a choice they did not make. More than this, those who do not notice the switch go on, quite fluently, to offer a justification for their supposed choice. These justifications do not differ in detail, or in cogency, from the justifications offered by the first group of subjects, those who are justifying choices they did, indeed, make. Subjects are not only unaware that experimenters have misled them about the choices they made. The justifications they offer for their supposed choices are offered with sincerity and conviction. They take themselves to be explaining the basis for a choice they made only moments before.¹⁰

It is tempting to explain away these results by supposing that the subjects were not paying attention; that they did not care about the choices presented to them; or that they were trying to please the experimenter rather than sincerely answer the questions posed to them. But all of these explanations of the phenomenon have been carefully controlled for and rejected. Subjects genuinely take themselves to be explaining why they chose as they did, even though the choices they offer explanations for are ones they did not actually make. The phenomenon turns out to be quite robust, and it builds on the important work of Richard Nisbett and Timothy Wilson done years earlier.¹¹

Nisbett and Wilson argued that, appearances to the contrary notwithstanding, we do not have direct introspective access to our mental processes. Although we seem to be able to tell, just by introspection, how we arrived at a belief or choice, the appearances

¹⁰Petter Johansson, Lars Hall, Sverker Sikstrom, Betty Tarning, and Andreas Lind, “How Something Can Be Said about Telling More than We Can Know,” *Consciousness and Cognition*, 15(2006), 673-692; Petter Johansson, Lars Hall, and Sverker Sikstrom, “From Change Blindness to Choice Blindness,” *Psychologia*, 51(2008), 142-155; Petter Johansson, Lars Hall, and Nick Chater, “Preference Change through Choice,” in R. Dolan and T. Sharot, eds., *Neuroscience of Preference and Choice*, Elsevier, 2011, 121-141. For a review of this literature and a discussion of its philosophical implications, see Hayley Webster, *Self-Knowledge, Choice Blindness, and Confabulation*, Ph.D. dissertation, University of Massachusetts, Amherst, 2019.

¹¹ “Telling More than We Can Know: Verbal Reports on Mental Processes,” *Psychological Review*, 84(1977), 231-259.

here are illusory.¹² In fact, our judgments about our mental processes are the product of elaborate subconscious inference in which we attempt to reconstruct what the process must have been by which we reached the belief, or the decision, which we did. Such processes of rational reconstruction are not all inaccurate. Indeed, they are often quite accurate. They are, however, very far from the mark when, for example, our original decision or belief was less than fully rational, or, as in the choice blindness cases, where we are asked to explain a belief or decision that we did not actually make. In effect, it seems that the manner in which we reconstruct our mental processes involves the implicit assumption that those processes were rationally based. As a result, our judgments about the origins of our beliefs are particularly misleading on precisely those occasions when we would like to be able to call upon reflection to locate our errors.¹³ The more we need correction, the more misleading our reflective judgments will be about the processes by which our beliefs and decisions were reached.

Sosa's very commonsensical suggestion that reflectively screened beliefs are more likely to be true than beliefs which are not reflectively examined turns out not to be correct. There is, I believe, an interesting feature of a great deal of philosophical thought about reflective self-examination from the time of Descartes to the present day. Epistemologists have often worried about the status of our beliefs. Do we really have reason to believe the things we do? A variety of skeptical arguments are presented which cast doubt on the status of our beliefs and this is then seen as providing a motivation for epistemological theorizing about the nature of justification and knowledge, as well as for a more careful reflective examination of our beliefs. The results of this reflective examination are then presented as a source of reassurance about the legitimacy of the beliefs which reflection certifies. When this kind of project is presented by someone with the philosophical skill and the literary talents of Descartes, it is very, very hard to resist, and the history of epistemology since the time of Descartes bears witness to the power of his vision. But Descartes combines an

¹² See also Alison Gopnik, "How We Know Our Minds: The Illusion of First-Person Knowledge of Intentionality," *Behavioral and Brain Sciences*, 16(1993), 1-15 and 90-101.

¹³ See my "Introspection and Misdirection," reprinted in my *Second Thoughts and the Epistemological Enterprise*, Cambridge University Press, 2019.

undue skepticism about the legitimacy of unreflectively arrived at belief with an overly optimistic view of the powers of reflective self-examination to set things right. The beliefs we arrive at unreflectively are not nearly as problematic as Descartes feared, and reflective self-examination is not nearly as useful in correcting our errors as Descartes supposed. If we are to understand the nature of justified belief and knowledge, the nature of epistemic responsibility and rationality, we will need to have a clearer idea of just how reflection works, and how unreflective belief works as well, than is to be found in much of the philosophical tradition following Descartes.

3. Some Recent Philosophical Thinking about Reflection and Action

Let us look at one exceptionally influential piece of work on the connection between reflection and action: Harry Frankfurt's "Freedom of the Will and the Concept of a Person."¹⁴ Frankfurt begins by noting the differences between adult human beings, on the one hand, and children and non-human animals, on the other, that we have already noted: While the beliefs of children and non-human animals are exclusively about the world around them, adults are also capable of having beliefs about their own beliefs. Similarly, adults, unlike children and non-human animals, are capable of having desires which are directed at their own desires. And this means that adults have a far more complicated psychology than do children and non-human animals.

As Frankfurt sees it, children and non-human animals are capable of free action: they may act in ways which are driven by their desires so that they are likely, in the ordinary course of events, to get what they want.¹⁵ They are not, however, capable of what Frankfurt refers to as freedom of the will.

The human adult capacity to reflect on our beliefs and desires allows us a degree of flexibility in our behavior of which other

¹⁴ Reprinted in his *The Importance of What We Care About: Philosophical Essays*, Cambridge University Press, 1988, 11-25.

¹⁵ Actually, it is slightly more complicated than this. They are likely to get what they want, on the assumption that their beliefs are true. That is, if their beliefs were true, the actions driven by their desires would serve to get them what they want. But this complication will not be important for the issues discussed in the text.

animals are incapable. If a child or a dog desires food, then, other things being equal, this will produce behavior which is designed to satisfy that desire. But if you or I want food, we may stop to reflect on whether we really want to act on that desire. I may have a strong desire for chocolate cake, but when I reflect on that desire, I would prefer not to act on it. Indeed, perhaps because I wish to stick to my diet, and having the craving for chocolate cake is so frustrating, I actually wish that I did not even have the desire for cake. As Frankfurt sees it, this possibility of surveying our desires and thinking about whether we actually wish to be moved by them, or whether we wish to have them at all, makes us different sorts of creatures altogether from young children and non-human animals. It is this possibility of either endorsing our desires and, in virtue of that endorsement, acting on them, or, alternatively, rejecting our desires, and, in virtue of that rejection, failing to act on them, that makes us capable of freedom of the will, something which children and non-human animals are, according to Frankfurt, incapable of. As Frankfurt describes it, a creature incapable of reflecting on its beliefs and desires is “a helpless bystander to the forces that move [it].”¹⁶ The capacity to reflect on one’s desires and deliberate about whether one wishes to be moved by them, however, transforms us from passive bystanders who merely witness our own behavior into genuine agents—persons, to use Frankfurt’s terminology—responsible for our own behavior.

Such a view accords nicely with the common sense account of the differences between adult human beings and children and non-human animals surveyed at the beginning of this paper. But there is a problem which this view faces. Frankfurt recognizes that my desire for chocolate cake may, in an important sense, not be fully my own. I may encounter it as an alien force operating within me, one I wish to resist. As Frankfurt sees it, the desire becomes fully my own when it is endorsed by a higher-order desire. Without that higher-order endorsement, the desire for chocolate cake is merely a force operating within me, something I may be witness to, but not something fully my own.

So suppose that I reflect on my desire for chocolate cake, and, unlike the self-controlled dieter, I have no need at all, and no desire, to diet. I not only desire chocolate cake, I like having this desire, and I fully enjoy acting on it. I thus form the desire to act on my

¹⁶ *Op. cit.*, 21.

desire for the piece of cake. I endorse my desire for the cake and, it seems, make it fully my own. By having this higher-order desire to act on my desire for the cake, my action of eating the cake manifests my freedom of the will.

Or does it? For consider now my higher-order desire: my desire to act on my desire for the cake. This desire was arrived at as a product of reflection. I reflected on my desire for the cake, and considered whether I really wished to be moved by it. But although I have reflected on my desire for the cake, I have not reflected on my desire to be moved by my desire for the cake. And, on Frankfurt's view, that makes my higher-order desire a mere force that is operating within me, rather than a desire which is truly my own. If this higher-order desire plays the crucial role in determining my behavior which Frankfurt has assigned it, I am then merely moved by a force within me, rather than moved by a desire which is wholly my own. To make it my own, I would have to reflect on this higher-order desire. But, of course, this would not solve the problem. It would just reproduce it at the higher-level.

This is, of course, the exact same problem which arose for Bonjour's view of justified belief. If beliefs which are not reflected upon are not justified, we cannot justify them by reflecting on our evidence for them, since the resulting beliefs about our evidence will themselves be unjustified, since they have not yet been reflected upon. Reflecting on them, however, will not solve the problem. It will simply reproduce the problem one level up.

It should be no surprise that Frankfurt's view runs into this problem. The idea that children and non-human animals are "merely pushed around by their desires," if that is a problem, cannot be solved by introducing still more desires, even higher-order desires such as the desire to be moved by particular desires.¹⁷ If the worry is that our mental states somehow push us around, being pushed around by more of them, and more complicated ones, is hardly reassuring. So either we should see the role which beliefs and desires play in determining action as presenting no impediment to freedom of any sort, or we should see the added complexity which higher-order desires afford us as making us no more free than other creatures who lack them. But either way, we

¹⁷This problem for Frankfurt was pointed out by Gary Watson in "Free Agency," *Journal of Philosophy*, 72(1975), 218.

should not see some fundamental difference in kind between creatures who have this kind of complexity and those who lack it.

Just as Frankfurt's view about freedom of the will faces the very same problem as Bonjour's view about justified belief, it also faces problems as a result of our current understanding of the psychological processes involved in reflection. Frankfurt addresses the problem of freedom of the will by examining the character of human action from the perspective of the actor him or herself. Frankfurt describes the deliberative process from the perspective of the deliberator, and he takes that perspective at face value. Thus, when the deliberator engages in some sort of interior monologue, thinking through the pros and cons of various courses of action, Frankfurt simply assumes that this monologue accurately reflects the decision process of the agent. But as Nomy Arpaly nicely points out, one's interior monologue need not accurately reflect the psychological forces at work in the deliberator. Someone thinking through the question of whether to eat a piece of chocolate cake may think, "I really enjoy chocolate, and that cake looks great. But I need to stick to my diet so I can lose the weight I've been putting on. I don't want to act on the basis of my desire for the cake. I'll just skip it." As Frankfurt presents it, this is the monologue of the self-possessed dieter, someone who acts on the basis of a higher-order desire not to be moved by the desire for cake. This is a paradigm case of freedom of the will, rather than letting oneself be pushed around by desires one just happens to have. Such a monologue may indeed be the product of a self-possessed agent, but it may equally be the monologue of someone suffering from anorexia, compulsively starving himself to death, all the while presenting his choice in his interior monologue as a model of rationality.¹⁸ One cannot take the perspective of the deliberating agent, whether about what to believe, or what to do, as an accurate record of the psychological processes which produce belief and action. The literature which we so briefly reviewed earlier, involving confirmation bias, confabulation, and choice blindness only serves to deepen this point.

¹⁸Nomy Arpaly, *Unprincipled Virtue: An Inquiry into Moral Agency*, Oxford University Press, 1996, 18.

4. Summing Up Thus Far

We have seen that the human capacity to reflect on our beliefs and desires is regarded by quite a number of philosophers, whether they are examining epistemological questions about our beliefs or questions about the nature of free action, as deeply revealing of important differences between adult human beings and children and other animals. Adult human beings are, to be sure, more complicated creatures, psychologically speaking. But these philosophers see the capacity to reflect on our mental states as making us importantly different in kind from children and other animals. In both the examination of epistemological issues and the examination of free action, this account of such important differences is aided and abetted by a methodology which takes the perspective of the deliberator at face value; it assumes that the deliberative process proceeds in just the way it seems to proceed when one engages in deliberation. But as I have argued, this methodology is fundamentally mistaken. An accurate account of the deliberative process shows that we are subject to an illusion when we deliberate. We seem to be directly aware of the psychological processes by which our beliefs and our actions are produced, and it is this illusion which has shaped a good deal of the philosophical discussion in epistemology and action theory. What is needed in these areas is a philosophical account which is informed by our best current understanding of the psychology of deliberation. As I have argued, there is every reason to believe that such an account will look very different from some of the most influential philosophical views currently on offer.

Just a few years ago, this is where I was content to leave the matter.¹⁹ As I saw it then, our perspective on our beliefs and choices, when we stop to reflect on them, is often quite inaccurate. It leaves out a great deal, so that we are typically unaware of many of the factors which influence our beliefs and our choices. And even as far as it goes, much of the picture it presents to us of our own decision process is mistaken, as the literature on confabulation shows. The process of deliberation, both about what to believe and about what to do, is far less efficacious in influencing our beliefs and choices than it seems to be when we reflect. When it is efficacious, it will often make us more confident in our beliefs and our

¹⁹ In *On Reflection*.

choices, but no more accurate or effective in our decisions. It can, to be sure, at times, improve our beliefs and our decisions, but it can also make our beliefs and choices less accurate and less effective.²⁰

I stand by all of these conclusions. But I now think that this understates the extent to which reflection on our beliefs and decisions distorts our view of our own mental processes. In all likelihood, I now believe, our mental processes are nothing like the way they appear to us when we deliberate. Let me explain the basis for my newfound radicalism.

5. What Belief Change and Decision Making are Really Like

For ease of exposition, let me focus on belief change, although my conclusions here apply just as much to decision making.

When we deliberate about what to believe, we seem to be able to survey our beliefs, or at least many of them, and assess the strength of our reasons for and against a claim that we wish to question. To return to my earlier example: If I find that I am tempted to believe something about the coming presidential election, I may stop to reflect on my evidence in order to figure out whether I should believe as I am tempted to. I try to figure out whether I have genuinely good reasons for this belief, or whether I might be engaging in wishful thinking. In trying to figure all of this out, I present myself with a variety of arguments designed to reveal the strength of my evidence. If I can construct a strong argument for some claim about the result of the coming election, from premises I reasonably believe, and I can find no convincing counter-argument that is similarly based, then I should follow where the evidence leads and endorse the conclusion of the argument I have constructed. This is what we seem to do when we deliberate.

Now the points we have already made about the psychology of reasoning show this picture to be badly oversimplified. There are premises we rely on without our being aware of them. And some of the premises we think we rely on are ones which play no role in our reasoning. But the picture I have presented thus far did not

²⁰ Jamin Halberstadt and Timothy Wilson, "Reflections on Conscious Reflection: Mechanisms of Impairment," in Jonathan Adler and Lance Rips, eds., *Reasoning: Studies of Human Inference and its Foundations*, Cambridge University Press, 2008, 548-565.

challenge the idea that the way to think about human reasoning is on the model of argumentation. When we reason, whether self-consciously or automatically, it is quite natural to assume that what goes on in us just is a matter of internal argumentation. When all is proceeding as it should, our beliefs are determined by good arguments from premises we properly accept.

This picture of human reasoning is very widely taken for granted in the literature on epistemology.²¹ Nevertheless, there is little attempt in this literature to show that this assumption squares with what we know from experimental work in psychology. This is unfortunate, because this is an area in which psychological work has a very direct bearing on how we should view epistemological issues.

One philosopher who has examined these issues in detail is Jerry Fodor. Fodor has defended the view that human thought is encoded in mentally represented sentences, although these sentences are part of a distinctive mental language—a language of thought—rather than in any natural language, such as English, Chinese, or Swahili.²² Our beliefs are just stored sentences in the language of thought, related to one another by various logical relations, and our mind is structured in such a way as to be sensitive to the logical relations among these sentences. One may think of the mind, on this view, as a computational device: our beliefs are written into the device’s memory, and the device is equipped with a program which, when new information is acquired, draws out the consequences of combining that information with the data already stored in memory. On this view, human inference is just argumentation in the language of thought.

The vast majority of Fodor’s attention has been on what he calls “input modules.” For example, on this view, there is a dedicated bit of mental software devoted to visual information processing. Just as one’s word processing program can do one job and

²¹ I challenged this picture in “Beyond Foundationalism and the Coherence Theory,” reprinted in my *A Naturalistic Epistemology: Selected Papers*, Oxford University Press, 2014, 17-31, but I now think that it is possible to fill in the details to this challenge in ways I did not see then. A different line of attack on this approach is presented by Anna-Sara Malmgren in “Availability, Goodness, and Argument Structure,” forthcoming.

²² *The Language of Thought*, Thomas Y. Crowell Company, 1975; *The Modularity of Mind*, MIT Press, 1983; *LOT 2: The Language of Thought Revisited*, Oxford University Press, 2008.

do it well, the visual information processing system takes visual input and transforms it into information which can then be made available to the rest of the mind, but it does nothing else. The visual processor is able to do what it does both quickly and reliably because it doesn't draw on all of the information we have stored in memory. It makes use of a small dedicated body of information which allows it to perform its job, and the fact that this body of information is insulated from the rest of what we know vastly simplifies the information-processing task it needs to perform: it doesn't need to sift through the large store of information in our memory, and thus it doesn't need to figure out which parts of that large store of information are relevant to its information-processing task. Each of our input systems works in this way, thereby making what would otherwise be an immensely complicated task into something tractable.

Of course, these input systems must eventually make their results available to what Fodor calls "central processing." And this is where the problems with Fodor's picture begin to emerge. Indeed, Fodor himself is well aware of these problems. As Fodor argues, we now know a great deal about various input systems. The fact that their information-processing tasks are narrowly defined, and that they each operate on the basis of a small body of information insulated from the rest of what goes on in cognition, is exactly what makes these systems easily amenable to study. But if we want to know about central systems, where belief acquisition and revision take place, no such restrictions apply. As Fodor makes clear, central systems must draw on all of the information we have stored in memory. The processes of belief acquisition and revision, as Fodor argues, is global, unlike the local information-processing that goes on in the various input modules.

And this leads Fodor to endorse what he self-mockingly calls Fodor's First Law of the Nonexistence of Cognitive Science: "the more global...a cognitive process is, the less anybody understands it."²³ As Fodor sees it, we don't understand anything at all about central systems, and thus, nothing at all about how belief acquisition and revision take place because we don't understand how the kinds of computational processes which Fodor builds in to the mind could possibly operate on the vast body of information which we have stored in memory.

²³ *The Modularity of Mind*, 107.

Sweeping as this conclusion is, I think it understates the implications of Fodor's approach. Fodor's picture of the mind as a computational device which stores information in sentences in a language of thought and which processes information based on the logical relations among those sentences does not just leave us with no understanding of how belief acquisition and revision work. Rather, this picture makes belief acquisition and revision impossible.

Let me explain. Fodor rightly sees that input systems, as he defines them, are able to do their job precisely because they have access only to a small body of information. Central systems, which have no such limitation, must work with all of the information which we have stored in memory, all of which, as Fodor sees it, is relevant to every single instance of belief acquisition and revision. When computer scientists try to define the difficulty of a computational task, they are interested in just how much computing time the task will take. And it is always the case that sifting through more information will take more time. But some tasks are such that the length of time they take grows in a linear fashion with the amount of additional data they need to work through: if they need to engage with twice as much data, they will need twice as long to complete the task; if they need to engage with three times as much data, they will need three times as long to complete the task; and so on. Other tasks are such that the length of time needed to complete them grows in an exponential fashion with the amount of additional data needed to be worked through: if they need to engage with twice as much data, they will need four times as long to complete the task; if they need to engage with three times as much data, they will need nine times as much time to complete the task; and so on. When computational tasks expand at an exponential rate, large data sets can make it impossible to complete the task.

As it turns out, some seemingly simple mental operations suffer from a combinatorial explosion: the amount of time it takes to complete them grows at an exponential rate as the amount of data they need to sift through increases. Checking a set of sentences for logical consistency is such an operation. As Christopher Cherniak has shown, something as apparently innocent as checking the consistency of a set of a few hundred sentences, given certain idealizing assumptions, would take from now until the heat death of the

universe.²⁴ And what this means, in short, is that no possible computational device can do this. And if no possible computational device can do this, then, if our minds work the way that Fodor suggests they do, then we cannot do this either. The problem here generalizes. The kinds of computations which a Fodorian mind would need to carry out in its central system simply cannot be done. Our thinking cannot be carried out by way of mental representations which correspond to good arguments carried out in natural languages. Since self-conscious deliberation gives us the impression that we are actually reasoning by way of such arguments, self-conscious deliberation misrepresents the nature of our thought. Our beliefs are not actually brought about by way of processes which are anything like the way they seem to us when we deliberate.

What then is the alternative? David Hume suggested that belief acquisition is entirely a matter of the association of ideas: when one kind of idea is regularly followed by another—for example, the idea of one billiard ball striking a second is regularly followed by the idea of the second ball moving away—we come to expect that such a pattern will continue in the future. As Hume saw it, this was not a matter of reasoning. We do not apply any sort of inferential rules in coming to our conclusions. Rather, the fact that ideas relevantly like the first are regularly followed by ideas relevantly like the second causes us to form certain expectations.²⁵ This Humean idea has been given contemporary expression in the work of various connectionists.²⁶ Of course, connectionist models of human thought do not at all resemble the sorts of reasoning that we seem to go through when we deliberate. But that is precisely the point. Our belief acquisition and revision cannot possibly work the way it seems to when we deliberate, so the fact that connectionist models have this consequence cannot be counted against them. Any work-

²⁴ *Minimal Rationality*, MIT Press, 1986, Chapter 4.

²⁵ *A Treatise of Human Nature*, L.A. Selby-Bigge, ed., Oxford University Press, 1967, originally published 1739; *Enquiries Concerning the Human Understanding and Concerning the Principles of Morals*, L.A. Selby-Bigge, ed., Oxford University Press, 1966, (originally published 1777).

²⁶For foundational work on connectionism, see James McClelland, David Rumelhart, and the PDP Research Group, *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*, vol 1: *Foundations*; vol. 2: *Psychological and Biological Models*, MIT Press, 1986.

able account of human cognition will inevitably have this consequence.

Connectionist models have one very simple rule of revision: an analogue of Hume's law of association. There are other models of cognition which import more structure than this, without going so far as the Fodorian view that cognition is a matter of argumentation. Philip Johnson-Laird has proposed an account of mental models which charts just such a middle course.²⁷ Problem solving is achieved, on this view, by the manipulation of mental models in rule-governed ways, but the models themselves do not correspond to sentences, and the mental manipulations do not correspond to anything like natural language argumentation.

There are other approaches as well. As Fodor rightly suggested, our current understanding of central systems—that is, of belief acquisition and revision—is in a primitive state. What we do know, however, is that these processes do not actually work in the way they seem to when we engage in deliberation. Deliberation gives us the illusion that we are forming and revising our beliefs by way of mental processes that track natural language argumentation. More than this, we seem to be directly aware of the processes by which our beliefs are formed when we deliberate, and this too is an illusion. It now seems that a proper understanding of cognition will show that how the mind actually works is not only different from how it appears to work when we stop to deliberate. It is not even roughly like how it seems to work when we deliberate. Just how it does work, however, remains to be seen.

6. Conclusion

Too much of epistemological theorizing has presupposed a common sense conception of how the mind actually works. Prior to the advent of modern experimental psychology, there was little alternative to such an approach. But incomplete as our contemporary understanding of our mental life is, we have made substantial advances since the psychology available to Descartes and Hume. Any epistemology worth its salt must engage with our best available theories, even as we recognize that progress in our psychological understanding will inevitably bring with it modifications in our epistemological views. The recognition that advances

²⁷ *Mental Models*, Harvard University Press, 1983.

in psychology will force such modifications does not license a methodology which simply turns its back on what we know.

Common sense ideas about the value of deliberation and how it provides a check on unreflective processes of belief acquisition and decision lead to views which valorize reflective thought over automatic processing. Such views are no better than the opposite extreme on which one is told never to reflect and simply to trust one's gut. The truth is, unsurprisingly, far more complicated than either of these two approaches would have it. Where reflection will benefit our thinking, and where it will lead us astray, where it will simply make us more confident, without making us more reliable believers and more effective decision makers, is something we cannot determine without very substantial experimental input. When we come to understand the process of reflection for what it truly is, rather than how it appears to us when we deliberate, we will better understand when reflection plays a constructive role in our thinking, and when it does not. What our current understanding seems to suggest, however, is that even where deliberation plays a constructive role, it works in ways wholly unlike the way it appears to us.

I will not be surprised if many see the view of the mind I present here as overly mechanistic. Viewing non-human animals in this way is completely untroubling. The idea that adult humans have minds which, while more complicated than those of non-human animals, are no less mechanistic in their operation, does not go down so easily. It will be said that other animals have minds which are moved by hard-wired mechanisms, while we actually reason. Their beliefs are produced in them passively, it will be said, while we bring about our beliefs. Other animals are not responsible for their beliefs or actions, it will be said, while it is an essential feature of human persons that we are responsible agents.²⁸ I have been arguing that all of these contrasts are vastly overstated and

²⁸ See, for example, these remarks by Paul Boghossian:

I don't see that a slight increase in the complexity of a hard-wired algorithm governing a creature's behavior can, all by itself, make for the difference between reasoning and mere mechanism, no matter how much the more complex algorithms might contribute to a 'more sensitive' interaction with the environment.

"Reasoning and Reflection: A Reply to Kornblith," *Analysis*, 76(2016), 46.

that they flow from a view of adult human minds which mystifies, rather than explains, our mental life.

When Darwin proposed the theory of evolution through natural selection, his approach was met with concerns that it somehow denigrated the human species. Darwin did not see it that way at all. “There is grandeur in this view of life,” he wrote.²⁹ The recognition that the human species is part of the natural world, and a product of the same forces responsible for the existence of other species, was not something that Darwin regarded as degrading. It is, instead, he thought, something we should look on with great awe and appreciation. Our evolutionary history is a remarkable thing, and an understanding of the extraordinarily complicated route by which our species came about is far preferable to any mythologizing or mystification that might be substituted for it.

I regard the mechanistic picture of the human mind in much the same way. This is not a degrading picture of human mental life. It is one we should look upon with respect and admiration. The human mental mechanism is a wondrous thing. A genuine understanding of how it works, in all its complexity, is something we should seek without fear that we will somehow be diminished in coming face to face with our place in the natural world.³⁰

²⁹ *On the Origin of Species*, Joseph Carroll, ed., Broadview Texts, 2003, (originally published 1859), 398.

³⁰ Thanks to Regina Kornblith for helpful comments on a draft of this paper, and to the audience at Brockport for a challenging and very helpful discussion.