[Manuscript version. Final version published in *Mind* and available at https://doi.org/10.1093/mind/fzz012]

As If: Idealization and Ideals, by Kwame Anthony Appiah. Cambridge, MA: Harvard University Press, 2017. Pp. xvi + 218

Adam Toon
University of Exeter
a.toon@exeter.ac.uk

Kwame Anthony Appiah's engaging and insightful new book focuses on idealisation. Based on three Carus Lectures delivered at the 2013 Eastern Division Meetings of the American Philosophical Association, the book takes its inspiration from the German philosopher Hans Vaihinger and his *The Philosophy of "As If"* (1911). Long neglected, Vaihinger's work has recently been revisited by philosophers of science interested in scientific modelling, most notably Arthur Fine. Vaihinger's own interests were much broader, however, taking in metaphysics, mathematics, ethics, law, theology and economics. Appiah's range is equally impressive and his discussion applies Vaihinger's ideas across a wide range of areas, including philosophy of mind, economics, moral and political philosophy. In doing so, Appiah offers a striking and extremely valuable insight into the pervasive role of idealisation in human thought.

As Appiah sees it, the central lessons of Vaihinger's work are twofold: "first, that in idealization, we build a picture—a model—of something that proceeds *as if* something we know is false were true; and second, that we do so because the resulting model is useful for some purpose" (p. 127). Idealisations are "useful untruths" (p. 1). In order to make sense of idealisation in a given domain, we must therefore ask what falsehoods are being treated as true and what purpose this is intended to serve. As Appiah's discussion shows, the answers can be many and various. Often our interest lies in controlling or

managing the world. To use a familiar example, a scientific model might proceed as if friction were absent from a system, in order to allow us to predict and control its behaviour. In other cases, we might instead be interested in "managing ourselves", as Appiah puts it. Thus, Vaihinger suggests that an atheist might still adopt religious claims for ethical and aesthetics purposes.

Appiah is keen to distinguish his project from more far-reaching anti-realist views. Such views are often labelled "fictionalist". For example, a moral fictionalist might argue that all moral claims are false and ought to be understood as acts of pretence. Of course, this might come as a surprise to those who make moral claims. By contrast, we are often well aware that our idealisations are false. Appiah uses this contrast to try to narrow his inquiry. As he puts it, "I am interested […] in cases where we (believe we) have a grip on the notion of truth and yet we have reason to go on using a theory that is, in some way or other, for some reason or other, not true". (pp. xvi). I wonder if these cases can be separated quite so easily, though. After all, as Appiah points out, Vaihinger is interested "in cases where the user of the fiction is aware, *or can be made aware*, that what she is thinking is not true" (p. 4; emphasis added). And fictionalists will often try to persuade us that, once we reflect more closely on our ordinary talk in some domain, we come to realise that it is not straightforwardly true, but instead serves some other purpose. Still, even if this divide is less clear cut than Appiah suggests, this need not prevent us from recognising the importance of the phenomena to which he and Vaihinger rightly draw our attention.

After he has introduced Vaihinger's framework, Appiah devotes much of the remainder of the book to detailed discussions of the role of idealisation in different domains. Chapters One and Two both concern idealisations involved in our notions of belief and desire. Chapter One discusses this theme in Daniel Dennett's work on the intentional stance, while Chapter Two considers the more technical notion of degrees of belief and desires found in rational choice models. Finally, Chapter Three provides a fascinating and compelling discussion of the role of idealisations in moral and political philosophy. Here, Appiah is rather more sceptical about the value of some idealisations, such as John

Rawls' ideal theory. As Appiah acknowledges, most of the book's more novel and controversial arguments are to be found in its contribution to these more specific debates, rather than in its overall remarks on idealisation. In what follows, I will focus on Appiah's discussion of idealisation in the attribution of beliefs and desires, but I hope this summary makes clear that the book contains much more that is of interest.

As mentioned already, Appiah's discussion of belief and desire centres on Dennett's work on the intentional stance. Although he acknowledges that Dennett might not agree with his interpretation, Appiah suggests that Dennett's views on intentionality "can be taken as a case study in Vaihinger's philosophy of the "as if"" (p. 34). In fact, Appiah writes, in Dennett's notion of the intentional stance, "what we seem to have is about as straightforward an application of Vaihinger's idea as you could get—because to adopt the intentional stance toward a person is to treat her *as if* she were a rational agent with beliefs and desires—the beliefs and desires "she ought to have given [her] place in the world and [her] purpose"—and then to predict what this rational agent will do in order to further her goals" (p.35). A little later on, Appiah puts the point slightly differently: "adopting a stance of this sort involves treating something *as if* something were so: as if it had internal states of belief and desire" (p. 36).

These characterisations of Dennett's ideas would seem to point towards two different sorts of as if thinking that might be identified in folk psychology, however. First, we might claim that folk psychology treats people *as if they were rational*. Second, we might claim that folk psychology involves treating people *as if they had certain internal states*, such as beliefs and desires. These different sorts of as if thinking target different sets of questions that we can ask about folk psychology. Along with Appiah, let us assume that folk psychology is a theory, albeit perhaps one that we grasp only implicitly. The first sort of as if thinking concerns the *form* that this theory takes. What principles govern our attributions of beliefs and desires? Do these principles somehow involve the assumption that people are rational? The second form of as if thinking concerns the *attitude* that we take towards folk psychology. Is folk psychology an attempt to describe what goes on inside our heads? Or do we take a different, and perhaps more cautious, attitude towards

it? Although distinct, these two sets of issues are related, of course. The attitude that we adopt towards a theory will typically depend upon the form that it takes.

It is the second sort of as if thinking—concerning the attitude that we take towards folk psychology—that is perhaps most commonly associated with Dennett's work on the intentional stance. Suppose that we understand folk psychology to be a theory about what goes on inside our heads. Broadly speaking, two options seem open to us. On the one hand, like Jerry Fodor, we might make an optimistic assessment of folk psychology's credentials and take it to be largely true. Let us call this *realism*. On the other hand, like Paul Churchland, we might be more pessimistic and claim that folk psychology is a false theory about our inner lives. Call this *eliminativism*. In this context, Dennett's ideas are often taken to offer an alternative both to realism and eliminativism, since he rejects their common starting point—namely, the idea that folk psychology aims to provide a theory of our inner states. Characterising Dennett's view more precisely is not straightforward, however, and Dennett himself has rejected many of the standard labels that come to mind.

One such label is *instrumentalism*. According to the instrumentalist, to say that someone has a particular belief or desire is not to make any claim about their inner machinery, but simply to say that it figures in the best predictive account of their behaviour. Notice that, strictly speaking, instrumentalism does not take folk psychology to involve any form of as if thinking, however. According to the instrumentalist, all there is to having a belief or desire is being a system whose behaviour is predicted by the intentional stance. Such a system straightforwardly *has* beliefs and desires; there is no "as if" involved. An alternative to instrumentalism is *mental fictionalism*. As I understand it, mental fictionalism claims that folk psychology treats people as if they had certain internal states, such as beliefs and desires, even if they do not (Toon 2016). Unlike instrumentalism, fictionalism acknowledges that it is part of our notion of belief and desire that they are internal states. However, the fictionalist claims that these internal states are useful fictions, not theoretical entities: we do not claim that people really have such states inside their heads; we merely pretend that they do. Although Dennett has objected to being called a fictionalist, such a view fits well with some of his remarks. For example, Dennett

compares beliefs to centres of gravity. Treating a system as if all its mass were concentrated at a point, even if it is not, can be useful for various purposes. Similarly, treating people as if they had inner states of belief or desire, even if they do not, can be useful for making sense of their behaviour.

Appiah's remark that the intentional stance involves treating a system "as if it had internal states of belief and desire" (p. 36) suggests that it is this second form of as if thinking that he has in mind. And yet his position on this aspect of Dennett's view is difficult to make out. At times, Appiah seems to take Dennett to be an instrumentalist. Thus, he refers, perhaps slightly warily, to Dennett's view as "part-time instrumentalism" (p. 51) and his characterisation of the intentional stance sometimes fits an instrumentalist reading. For example, he writes that "[w]hat it is, finally, to have beliefs and desires is to be an "intentional system, a system whose behaviour is reliably and voluminously predictable via the intentional strategy" (p. 37). On the other hand, Appiah is willing to concede that Churchland might be right to say that we do not really have beliefs and desires. And yet instrumentalism cannot allow for the possibility of eliminativism: according to the instrumentalist, if a system is predictable using the intentional stance, then it has beliefs and desires. Any discoveries that future cognitive science might make about its internal organisation are irrelevant to its status as a true believer.

Appiah rightly points out that, even if Churchland is correct that beliefs and desires do not exist, it is hard to see how we could avoid talking about them. As I have noted already, fictionalism offers one way to make sense of this: we might continue to talk as if people had such internal states, even if they do not. However, much of what Appiah says suggests that fictionalism is not the sort of position he has in mind. In fact, much of his discussion points towards realism. For example, he asks why the intentional stance works and says "if I have beliefs and desires and am rational, the reason the intentional strategy of treating me as a rational agent works, when it does, is: that I am a rational agent with those beliefs and desires" (p. 38). This seems to require that we adopt a form of realism. Certainly, that we possess beliefs and desires in the instrumentalists' sense cannot explain the success of the intentional strategy, for this would amount to saying that the strategy

works because it works. Furthermore, in his discussion of decision theory in Chapter Two, Appiah explicitly adopts a functionalist and representationalist theory of mental states that stands at odds with both instrumentalism and fictionalism. Of course, Appiah is careful to note that the notion of degrees of belief is not itself part of common sense talk about the mind. And yet his analysis is intended as an amendment to folk psychology, rather than a radical revision of our ordinary notions of belief and desire.

Perhaps the reason that Appiah's views on this aspect of the intentional stance are a little unclear is that his main focus is on the first form of as if thinking involved in the intentional stance—namely, thinking of people as if they were rational, rather than as if they had certain internal states. According to Appiah, "the sort of rationality in question is extremely demanding: it involves having all the beliefs and desires we ought to have and acting only as we ought to act, given them" (p. 40). Unfortunately, people typically fail to live up to this exacting standard. Appiah offers the example of "birthers", who refuse to believe that Barack Obama was born in Hawaii, and smokers, whose desire to smoke might seem far from rational. The upshot is that "[i]f you have to be fully rational to have beliefs and desires, then I don't have beliefs and desires and neither (excuse me for saying this) do you" (p. 40). Only a being that was fully rational—which Appiah calls a *Cognitive Angel*—would have fully fledged beliefs and desires. The rest of us must make do with what Appiah (following Dennett) calls "sorta" beliefs or desires: "[i]n the actual world […] every belief is a "belief"—a sorta belief—and every desire is a "desire" (p. 43). The result is a form of eliminativism: "the right answer to the question whether anything at all really has a mind can be: sorta. But being sorta true is not, alas, a way of being true—it is a special way of [being] false" (p. 45). Notice that this eliminativism is seemingly even more radical than Churchland's, for it implies that we do not possess beliefs and desires even in the instrumentalist's sense: if the intentional stance demands ideal rationality, then we are not even intentional systems, never mind what future cognitive science finds inside our heads.

Appiah's decision to focus on this aspect of the intentional stance makes sense given the way in which he characterises his overall project in the book. As we saw, Appiah aims to

consider cases in which, although we have a "grip" on the truth, we nevertheless stick with claims we know to be false. The sort of fictionalism about mental states that I outlined above—which claims that folk psychology treats people as if they had certain internal states—seems at odds with this characterisation. For surely, it might be argued, the folk do not *know* that people do not have the required internal organisation. Indeed, part of the intuitive appeal of fictionalism and related approaches to the mind, such as instrumentalism, rests on the idea that folk discourse is simply not concerned with our inner workings. On the other hand, the first form of as if thinking—thinking of people as if they were rational—fits Appiah's characterisation more comfortably. For the folk are all too well aware that people will typically fail to meet the required standard of rationality. And yet putting matters this way makes this aspect of Appiah's discussion seem more puzzling. Why should we think that our ordinary attributions of beliefs and desires invoke this exacting standard of rationality when, as Appiah readily admits, we know perfectly well that people fail to meet it?

Dennett himself has received considerable criticism on this score. Many are willing to grant that folk psychology assumes some minimal standards of rationality: if someone's behaviour is truly bizarre then it can be hard to know what they want or believe. But the idea that our ordinary attributions of mental states invoke the more stringent standards of rationality enjoyed by Cognitive Angels is hard to maintain. Consider some examples from Shaun Nichols and Stephen Stich (2003). Suppose that John knows he is allergic to chocolate but still eagerly unwraps and eats a bar in front of us. John's behaviour is hardly rational. And yet surely John *wants* to eat the chocolate. Our folk psychological practices would still lead us to attribute this desire to John—although we might well say that his desire was rather unwise. Or consider Kahneman and Tversky's famous experiment involving the "feminist bank teller", Linda. Given some information about Linda's background (e.g. that she has taken part in anti-nuclear demonstrations), most subjects judge that the statement (a) "Linda is a bank teller and is active in the feminist movement" is more probable than the statement (b) "Linda is a bank teller". Committing this "conjunction fallacy" is irrational. And yet surely the subjects in the experiment

believe that (a) is more likely than (b). Indeed, it can be hard to persuade them otherwise!

Examples like these would appear to show that folk psychology is perfectly comfortable attributing beliefs and desires despite our well-known failures of rationality. This suggests one response to the threat of eliminativism that Appiah's discussion raises. Rather than concluding that we do not have beliefs and desires, we might instead decide that we have simply mischaracterised the principles of folk psychology. However, Appiah himself suggests an alternative, and intriguing, response to such concerns. This response draws on Nancy Cartwright's work on models in physics. According to Cartwright, even if a model is strictly false, it might nevertheless reveal important truths about the underlying capacities that operate within nature. For example, a model of a falling object might reveal something about gravitation, even if it is strictly false due to inference by wind, air resistance, and so on. Similarly, Appiah suggests, the intentional stance might be seen as an idealised model that reveals an underlying truth about beliefs and desires, even if it is strictly false due to interference from other factors. In this way, we might even recover a form of realism about mental states. As Appiah puts it, on this view, "[w]e really have beliefs and desires. They would work as in a Cognitive Angel if there were no other forces operating in our minds to get in the way. The idealization is Galilean: it is supposing—acting as if—there are no other forces." (p. 48)

This is an ingenious proposal. However, I wonder if the parallel that Appiah draws might be misleading in certain respects. First, notice that, on Cartwright's picture, fundamental laws are, first and foremost, characterisations of the behaviour of idealised models, while real systems satisfy these laws imperfectly at best. Our theories thus give us fairly immediate epistemic access to the behaviour of our models; applying these models to the world proves more problematic. Likewise, Appiah's analysis suggests that the principles of folk psychology are, first and foremost, a characterisation of ideally rational agents. It is Cognitive Angels who satisfy the principles of folk psychology and in whom beliefs and desires run their proper course. Limited, fallible creatures like human beings satisfy those principles imperfectly at best, so that applying folk psychology here is likely to prove more problematic. And yet this seems to put matters the wrong way around. On the

face of it at least, our ordinary talk about the mind has its proper home in our interactions with other limited, fallible creatures—that is, with each other—rather than in the characterisation of ideally rational agents. Personally, I find it much easier to make sense of the behaviour of other, somewhat-irrational creatures like myself than I do to imagine how a Cognitive Angel—or (to use some of Appiah's other examples) Star Trek's Mr Spock or the android Data—would think and behave. One way to put this point is to note that it is typically easier to say what someone *does* believe than what they *ought* to believe. If a friend tells me they are a Christian, I immediately assume that they believe in God. I find it much more difficult to decide whether someone should believe in God.

Second, Appiah's analysis also suggests that, in order to explain cases of irrationality, we must identify other factors that interfere with the normal functioning of belief and desire. We do not worry if our model of the falling object gives poor predictions, as long as we can point to other factors that are responsible, like wind and air resistance. Appiah draws a similar lesson in his discussion of decision theory: "if the agent's behaviour deviates from what the theory requires, this must be the result of an *independently specifiable causal intervention* with her mental functioning" (p. 80). At least in the case of folk psychology, however, this demand seems too strong. It is true that we explain some apparently irrational behaviour in this way. Appiah himself gives the example of paralysis, where a difficulty with someone's muscles might prevent them from acting as our theory of rationality might expect. But not all deviations from ideal rationality are explained in this way. Sometimes, we might find ourselves simply at a loss to explain why someone holds irrational beliefs or desires. At other times, our explanations might lie entirely within the domain of folk psychology itself. For example, Tversky and Kahneman explain their subjects' susceptibility to the conjunction fallacy by invoking what they call the *representativeness heuristic*: the subjects mistakenly judge it to be more likely that Linda is feminist bank teller than simply a bank teller since they judge her to be a closer fit to their idea of a typical feminist bank teller than their idea of a typical bank teller. Unlike the example of paralysis, this explanation of people's deviation from ideal rationality seems to operate entirely within folk psychology, rather than

invoking interference by other causal factors from outside.

Although I have focused on Appiah's discussion of belief and desire, and raised some concerns about his analysis of these notions, I hope that this review gives some idea of the wealth of insights to be gained from this concise and elegantly written book. Readers who are more familiar with debates over the role of idealisation in economics or moral and political philosophy will no doubt find Appiah's discussions of these areas equally thought-provoking. However, as I indicated earlier, it is the book's broad scope that is perhaps its main strength. In the Preface, Appiah notes that, "often in philosophy it is useful to stand back and take a broad view of a topic, knowing that real progress requires work with a narrower focus as well" (p. ix). The book's aim is thus "not so much to announce any startling discoveries as to persuade you that idealization matters in all the major areas of the humanities and the sciences and in everyday life, and to commend it as a topic of reflection and research (p. x). In this central aim—as well as in its more detailed discussions of particular domains—the book succeeds admirably.

References

Nichols, S. & Stich, S. (2003). *Mindreading: An Integrated Account of Pretence, Self-Awareness, and Understanding Other Minds*. OUP.

Toon, A. (2016). Fictionalism and the folk. *The Monist*, 99, 280-295.