

Cognitive error as absolutisation

Robert M. Ellis

Abstract

A unified explanation is needed of cognitive biases together with the fallacies identified in philosophical tradition, which also supports practical responses to cognitive error. If we focus on the normative element of cognitive biases and fallacies, the most adequate explanation is that of absolutisation of assumption, where absolutisation is the psychological state in which a belief is held with repression of alternatives. This approach can also incorporate the dialectical aspects of some accounts of fallacy, and allow for the variability of ends rather than relying on means-end rationality. This approach also provides psychological support for the thesis that metaphysical belief (as ontology rather than as general assumption) is a form of cognitive error because of its exclusion of alternative possibilities.

Introduction

A huge wealth of material has now been developed in cognitive psychology on the identification of hundreds of cognitive biases, all of which reveal common patterns of error in universal human operation. However, due to the barriers between psychology and philosophy, little exploration seems to have taken place as to the philosophical implications of these patterns of error, or to establish their relationship with the fallacies identified by philosophy. The psychologists also seem to have left an unresolved area around the question of how we should respond to cognitive biases: one that depends on issues of responsibility. How far should we regard cognitive biases as avoidable or as unavoidable errors? This again is an area requiring philosophical exploration.

In this paper I want to put forward a thesis that addresses both these issues in cognitive biases. Firstly I want to try to differentiate those aspects of cognitive bias for which we may be responsible, and in relation to which cognitive biases may thus offer, not only a basis of therapy, but a moral and epistemological working ground for individuals. How far are cognitive biases descriptions of unavoidable features of human functioning, and how far within a sphere of responsibility? Secondly, I want to put forward a thesis about what our sphere of responsibility in cognitive biases may have in common with fallacies as identified in philosophy. My thesis is that this commonality is absoluteness of assumption. This second point will require first of all an argument that we should respond to cognitive biases in terms of assumption rather than in term of other possible conceptions of rationality, then also the presentation of evidence, at least in relation to some exemplar biases, that these assumptions are those of absolutisation.

I shall argue that this account of cognitive biases makes it easier both to find common patterns between biases so that we can respond to them practically (beyond the context of psychotherapy, as well as within it), and to make it clearer what that response might be. It will make it possible to describe, in general terms,

some ways of addressing (not wholly avoiding, but addressing) cognitive biases in ethics, politics, education and many other possible contexts.

A justification for applying considerations of responsibility to cognitive biases

The question of responsibility appears to be the elephant in the room in the discussion of cognitive biases. Though there is agreement that we can do something about the errors that we are prone to making, the extent of our responsibility for making them or not making them does not seem to have been generally defined by the psychologists commenting on these issues.

I will quote Daniel Kahneman at some length, as representing a highly informed and experienced position based on a lifetime's research of cognitive biases.

“What can be done about biases? How can we improve judgments and decisions, both our own and those of the institutions we serve and that serve us? The short answer is that little can be achieved without considerable expenditure of effort. As I know from experience, System 1 [fast, intuitive thinking] is not readily educable. Except for some effects that I attribute mostly to age, my intuitive thinking is just as prone to overconfidence, extreme predictions, and the planning fallacy as it was before I made a study of these issues. I have improved only in my ability to recognise situations in which errors are likely: “This number will be an anchor...,” “The decision could change if the problem is reframed...” And I have made much more progress in recognising the errors of others than my own.

The way to block errors that originate in System 1 is simple in principle: recognise the signs that you are in a cognitive minefield, slow down, and ask for reinforcement from System 2 [slow thinking]... Unfortunately, this sensible procedure is least likely to be applied when it is needed most. We would all like to have a warning bell that rings loudly whenever we are about to make a serious error, but no such bell is available, and cognitive illusions are generally more difficult to recognise than perceptual illusions. The voice of reason may be much fainter than the loud and clear voice of an erroneous intuition, and questioning your intuitions is unpleasant when you face the stress of a big decision.”
(Kahneman 2011, 417)

One striking aspect of this passage is that if we were to substitute moral responsibility for cognitive bias, the picture would be hardly likely to change. Imagine if Kahneman was an alcoholic trying to engage with addiction, and was reflecting on the difficulties of becoming aware of and changing this moral behaviour. Perhaps, through reflection and training, he could become more aware of the situations in which the problem responses were likely to occur, but nevertheless he might find ‘the voice of reason’ similarly drowned out by a stronger and more habitual intuitive response etched deeply into his neural pathways. Nevertheless, the recognition of moral responsibility would be valuable in these circumstances, whether in terms of individual practice, social systems of reinforcement or legal penalties. Why should cognitive biases be treated differently?

There are several common philosophical arguments as to why responsibility should not be applied to cognition, all of them dubious. I will consider two salient ones here. One may involve the separation of a moral faculty from other kinds of decision-making, and the insistence that cognitive judgments are non-moral. Another may involve an assumption of determinism coupled with moral conventionalism

The argument for a separate moral faculty (e.g. Hauser 2006) involves the continuation of a tradition of reasoning about ethics that goes back to Kant (1785/1959). Here moral judgments are limited to those that involve categorical imperatives, following or neglecting universalisable principles. Judgments that do not involve such imperatives at all (such as a judgment whether to brush your teeth in the morning) are believed to be morally neutral, merely 'prudential' judgments, even if they may sometimes appear to have the force of morality. Hauser attributes the moral faculty that can make such judgments to a special and unconscious psychological faculty, developed in the course of human evolution. However, Johnson (2014, 147-8) uncovers the vacuousness of this position, by showing what a wide range of psychological functions (including patience, categorization, empathy, and a narrative sense of self) would have to be included in this moral faculty even on Hauser's account of it. Kahneman's account of the skills needed to begin to engage with cognitive biases greatly overlaps with any such account of the features of a moral faculty – whether these features are those of self-awareness, patience (required in 'slowing down'), or awareness of our categorizations.

Even if there were such a thing as a separate moral faculty distinguishable from other psychological functions, this would not necessarily indicate that the notion of responsibility should not be applied beyond the sphere of that supposed faculty. Even if my decision to brush my teeth every day is not a moral decision in any sense, I could still be judged responsible (and hold myself responsible) for such a decision on prudential grounds, and be justified in doing so.

An alternative line of rejection of the notion of responsibility to cognitive biases is deterministic. A particularly pessimistic reading of the difficulties such as Kahneman reports in changing our biases may be that we *cannot* do anything about them, and that cognitive biases are merely descriptions of inevitably erroneous processes in human reasoning. At best, then, those who think along these lines will offer a compatibilist account of responsibility, in which responsibility is merely a determined social mechanism by which society exerts power over its members, through praise of socially beneficial behavior and condemnation of socially unhelpful behavior.

In response to this approach it can first be observed that determinism itself involves a metaphysical assumption beyond all observation. That by itself might offer sufficient reason for refraining from applying it, especially to issues where the extent of causal determination remains doubtful due to the complexity of the processes involved. However, as I am aware that determinism is a widespread prejudice, let me also offer an alternative response that would also apply here even if determinism were in some sense "true".

On the compatibilist account, the social function of responsibility remains unaffected by determinism. Thus, on this account, even if my belief that I can change my cognitive biases is in some sense illusory, perhaps itself driven by some further bias, this belief may nevertheless be effective in changing my behavior in beneficial ways. Although I do not accept the determinism that informs such a compatibilist account, I do agree with it that we could bracket out all questions about the ultimate causation of our moral responses. *Whether or not* they are ultimately determined, we can note that we feel responsible for some kinds of judgment, that we hold ourselves responsible for them, and that others may praise or blame us (or even reward or punish us) for those judgments we feel responsible for.

Let considerations of the ultimate origins of our sense of responsibility thus be no barrier to a phenomenological recognition that it is present, and that it can be applied, not just to issues commonly identified as ethical, but also to cognitive biases.

The extent of our responsibility for cognitive biases

Having sidelined such objections, then, we can then get onto the more interesting business of assessing *how far* responsibility applies to cognitive biases. There are clearly two extremes to be avoided here: one is the deterministic assumption that we have no responsibility, and the other the libertarian assumption that we could completely dispense with biases through a mere act of will.

The quotation from Kahneman given above already gives us grounds for arguing that we can do *something* about cognitive biases, even if such action involves many difficulties. Kahneman notes that he has become more likely to identify situations in which error is likely. He also does not appear to rule out the possibility that the “warning bell” or the “voice of reason” may come to our attention and help us avoid error, even if this only happens in a very small minority of cases. Changing our judgments by planning ahead and changing the conditions in which the judgment is made is not intrinsically different in terms of responsibility from making a different judgment without such preparation, merely more difficult. By analogy, if someone carries a fragile object carelessly and then breaks it, we are usually inclined to nevertheless hold them responsible for the breakage. In the same way, it can be argued that if we make cognitive judgments carelessly, because we have failed to prepare or train ourselves in advance in ways that might have improved those cognitive judgments, we are nevertheless responsible (at least to a small extent) for those judgments.

The basic mechanism by which an error of any kind may be noticed seems to be that of the attention developed in the medial pre-frontal cortex. An increasing amount of evidence shows the effectiveness of mindfulness training in developing such wider awareness of cognitive error (Siegel 2007, Jha et al 2007). At the same time, attempts to avoid cognitive biases through prior preparation and training are in any case already a feature of good scientific practice. For example, double-blind trialing in medical research has the goal of avoiding confirmation bias. Having been trained in the scientific procedures of double-blind trialing, a medical researcher who failed to use it when possible and appropriate would be

held responsible for doing so. Our responsibility for cognitive biases, then, may operate through the medium of institutions that are set up to help us avoid them, or may be a matter of more immediate personal awareness.

On the other hand, the limitations of our ability to change cognitive biases are highly evident in the ways those biases are related to highly entrenched, often genetically-determined, responses. Often biases are the necessary result of having a single embodied point of view, or they may be subject to evolutionary explanation as outcomes of early human adaptation. Let us take the example of anchoring. This is the tendency to begin our estimates about a new state of affairs on the basis of previous representations, rather than beginning afresh on entirely new evidence. Kahneman and Tversky demonstrated this by getting experimental subjects to produce random numbers on a wheel of fortune, then guessing the proportion of African nations at the UN (Kahneman 2011, 119): they found that the estimates were strongly influenced by the random number produced, which “anchored” the responses. Clearly there is much anchoring that we can do relatively little about given our neurological functioning, as once certain neural connections have been activated, they are difficult to discount or ignore until that activation has died down. Some degree of anchoring might well be inevitable, simply because the brain is not a *tabula rasa* each time it has a new thought, but something more like an oil painting where old brush marks remain until they are over-painted (and even then the underlying color may still influence that placed on top of it).

Thus it can be concluded that we have some degree of responsibility for any given cognitive bias, which in any given case is neither zero nor absolute, but somewhere in between. Of course, no general formula can be offered for our degree of responsibility beyond the mere avoidance of these extremes, because our degree of responsibility for a particular bias will depend not only on the nature of the bias, but also the circumstances and previous conditioning of an individual brain. Responsibility is, in any case, not a quantity that can be precisely measured; and it is generally more useful, once we have acknowledged a degree of responsibility, to focus not on precisely measuring it so much as on ascertaining the conditions we need to work with in order to fulfill that responsibility to any extent.

To do that requires me to now turn to a more general theory about cognitive biases and fallacies. However, I will do so now able to apply a conceptual distinction derived from this discussion of responsibility: namely that between the elements of a cognitive bias for which we are responsible and the elements for which we are not. I will call the elements of a cognitive bias for which we are responsible the *normative element* and those elements for which we are not responsible the *descriptive element*. In practice I do not think these two elements can be readily distinguished in any particular case, and they are certainly interdependent, but this is nevertheless a useful conceptual distinction. In this it follows the more general fact-value distinction on which it is based, which can be used conceptually, provided that when applying it we do not forget the impossibility of distinguishing it in practice.

Cognitive biases and fallacies as assumptions

Cognitive biases and fallacies seem to be identical or similar phenomena investigated by different methods. The notion of a fallacy within the philosophical tradition depends only on the conceptual identification of a type of error that *can* be made, whilst the psychological tradition lays stress on consistency of evidence, usually derived through experiment, showing the universality of tendency to a particular type of error. Fallacies are also often seen as features of explicit argument, whereas the fallacious 'arguments' detected by cognitive biases are usually a matter of implicit assumption: but since both types of error are usually unconscious and both can be explicitly analyzed, this is hardly a strong distinction. Another distinction is that fallacies are clearly normative: the point of identifying an error in philosophical terms is to encourage us to avoid it. In psychology, however, the normative and descriptive elements of cognitive biases are inter-related.

However, within each tradition, the definitions of both cognitive biases and fallacies are each highly contested. It is striking how similar the contests are in each field, each depending on the highly contestable notion of rationality. In the light of the similarities between the contests within both psychology and philosophy, the insistence that cognitive biases and fallacies must be intrinsically different phenomena (rather than similar phenomena that have been investigated in differing ways) seems entirely spurious. The question as to what makes a fallacy or cognitive bias an 'error' is the same question on both sides of the disciplinary division, and on both sides the most commonly given answers are equally unsatisfactory.

There are broadly four different accounts of error found in the discussion of both fallacies and cognitive biases:

1. Error as a wrong application of logic
2. Error as false belief
3. Error as deviant view, beyond social consensus
4. Error as failure of means-end rationality

I will argue in each case that such an account deals inadequately with the nature of the errors identified both philosophically and psychologically.

1. Error as a wrong application of logic

The most common and basic understanding of a fallacy is that determined by the model offered in formal fallacy (usually understood as a fallacy that can be seen in formal terms alone regardless of content). In a formal fallacy, a conclusion is drawn that is not necessarily true given the truth of the premises: for example, affirming the consequent (if P then Q, Q, therefore P). Although this pattern of invalidity does not always apply to informal fallacies, which may be formally valid, it may be argued that invalidity is still the core reason for fallacy. In the case of inductive argument, the equivalent of invalidity may be claimed to be weak support for an inductive conclusion.

Cognitive biases are also often assumed to be problems of logical processing. Perhaps the clearest example of this is the belief bias (Evans et al 1983). Evans and his associates found subjects working through a kind of sham reasoning

process which substituted an easier problem for a harder one, in order not to have to challenge a belief to which they were attached. Mistakes in estimating probability are also easily attributable to logical processing errors: for example the conjunction fallacy, in which the conjunction of two associated conditions is judged more probable than one of those conditions alone – Linda the philosophy major is assumed more likely to be a feminist bank teller than just a bank teller. (Kahneman 2011, 158).

At one level, this account of error is inadequate because it does not apply at all to the vast majority of informal fallacies and cognitive biases. That is because the vast majority of reasoning neither is, nor needs to be, strictly deductive in form, and thus its validity is simply not a relevant criterion to apply in assessing its justification. In fact, the attribution of a deductive pattern in an inductive context forms part of the error. For example, the fallacy of attributing a single cause to an event that may be multi-causal (such as automatically blaming the government for an economic crisis) involves the deduction of a single cause from the evidence of an effect (and perhaps other comparable events), rather than recognising the reasoning involved as inductive and thus imperfect and approximate.

However, this inadequacy also applies to formal fallacies and cognitive biases that seem to involve mistakes of logical validity. That is because we by no means have no choice but to interpret such mistakes only in terms of logical error. Every invalid conclusion is invalid because of the assumption of hidden and unacknowledged premises that are implied in drawing that conclusion. In the case of affirming the consequent, for example, we assume that 'if P then Q' also implies 'if Q then P'. If we include this assumption in the fallacy of affirming the consequent, we get a valid conclusion, showing that this assumption is responsible for the formal fallacy:

If P then Q

('If P then Q' also entails 'if Q then P')

Q

Therefore P

In the case of the belief bias, too, when we substitute sham for genuine reasoning to justify our preferred conclusion, we are assuming that this conclusion must be correct regardless of the reasoning. In the conjunction fallacy, we are assuming that the conjunction of conditions that is more psychologically available to us is more probable because of that availability.

If, then, we choose to focus on our reasons for making an apparent mistake in reasoning rather than assuming that this mistake must be conceptualized in terms of logical error, we find that logical error is neither sufficient in explaining all cases, nor necessary in accounting for error in any given case.

2. Error as false belief

A common alternative to the explanation of fallacies and cognitive biases in terms of logic is to accept that it is their premises that are often the basis of error. However, it is then common practice to describe such premises as 'false'. For example, we may explain the conjunction fallacy by claiming that subjects

assume *falsely* that the conjunction of two conditions can be more probable than one of those conditions alone.

However, the assumption that an unjustified assumption is necessarily false itself involves a fallacy. The subjects involved lack adequate justification for the conclusions they are drawing, but that only means that they do not know the conclusion to be true, not that it can be judged false. To assume otherwise involves a false dilemma. For example, if Peter makes a hasty generalization by assuming that his new Italian acquaintance, Giovanni, must like pasta because he is Italian, the problem is not that it is false that Giovanni likes pasta, but that Peter does not know *whether or not* Giovanni likes pasta.

If the supposed beliefs involved are not known to be false, then their falsity is practically irrelevant and will not help us to identify specific cases of fallacy or cognitive bias. If they are only 'false' in terms of *a priori* principles such as laws of logic or axioms of probability, then this involves an appeal to logic, and the arguments above apply. A third alternative is that they are just widely acknowledged or believed to be false, in which case we should categorize the error not as one of falsity but as one of social deviance.

3. Error as social deviance

Cognitive biases and fallacies can be seen as social deviance in two overlapping ways. One is to focus on assumptions made in a fallacious argument as contrary to 'truth' as generally recognized in one's context. Another is to focus on fallacious argument as employing strategies that involve breaking socially acceptable rules of argument. However, since reasoning strategies can always be understood alternatively as assumptions (as discussed above), these two ways of understanding cognitive error as social deviance do not really differ.

One persuasive account of informal fallacy along these lines is that of Douglas Walton (1987, 19ff.). Walton points out that in practice, the vast majority of arguments do not take the *monolectical* form of an isolated sequence of propositions supporting a conclusion, but actually have a wider social context that is *dialectical*, in the sense that they aim to persuade those holding a different view and depend on engaging the commitments of the other side. Even if those holding an opposing view are not physically present, they nevertheless form a context of discourse within which the argument needs to be understood. This paper would be an example that would fit Walton's account of appearing monolectical, but actually aiming to persuade people of one thesis rather than another when interpreted in a wider social context, because it aims to show how the widely accepted commitments of its likely readers provide support for the position I am putting forward. Here, Walton suggests, the rules of argument are those that both parties accept, and are thus contextually dependent. An informal fallacy, then, is an assumption used in the argument that breaches those implicitly agreed, but contextual, rules. For example, the distinction between an unfair *ad hominem* attack and a slightly pokey jocular remark is hugely contextual, and depends not just on features of a person being used to justify a conclusion, but also on the emotional atmosphere and context of interpretation.

Though at first sight Walton's account of error applies only to informal fallacies and not to cognitive biases, the more basic pragmatic function of the avoidance of error that he uncovers can be applied just as much to cognitive biases. A dialectical procedure overcomes conflict by allowing the grounds of agreement to be identified, so in terms of the rules of argument, the implication of Walton's approach must be that fallacies have the function of interfering in that process of agreement and perpetuating conflict. This is a model that could be applied to internal conflicts as well as those in external debate, with cognitive biases being seen as interferences in the ways that conflict between perspectives held by one person at different times can be resolved. I will be expanding on this approach below.

However, it is the relativist implications of an account of fallacy or cognitive bias as social deviance that create the most difficulties. If these errors are breaches of rules that are accepted in one context rather than another, our understanding of them is apparently deprived of normative power outside that context. Not only the precise application of a rule such as avoiding *ad hominem*, but even the rule itself may be seen as contextually dependent and thus relative. My suggestion is thus that we adopt the dialectical form of Walton's account in explaining fallacies and cognitive biases, but avoid its potentially relativist implication by exploring ways that it could be universally applied: which I will attempt to do below.

4. Error as a failure of means-ends rationality

Kahneman has had considerable success in undermining the assumption of universal means-ends rationality found in classical economics, and demonstrating that, due to cognitive biases, economic actors rarely behave in the ways predicted by that system. Certain cognitive biases are clearly amenable to the interpretation that they consist in a self-sabotaging tendency to prevent ourselves getting what we want: for example, the planning fallacy (Kahneman 2011, 249-251) involves a tendency to underestimate the costs and delivery times of projects, undermining the chances of bringing the project to fruition.

However, to explain all cognitive biases in terms of means-end rationality itself involves fallacious thinking. It assumes that our ends remain stable whilst we adopt different means (some more effective, and others, such as those involving biases, less so) to fulfill them. A number of cognitive biases reveal this picture to be an over-simplification, because they chart constant changes in our ends. For example, the hedonic treadmill describes the tendency of our levels of happiness to normalize once a desire has been fulfilled, with even major events like promotions or lottery wins making little difference to our levels of happiness (Fujita & Diener 2005). The significance of the end thus clearly seems to change once it has been achieved. The 'winner's curse' provides a further example showing that the value of ends is not stable: once we have won something against competition, we may well find ourselves disappointed with a result that does not match the exaggerated expectations we have built up whilst competing for it.

Rather than ends providing a stable basis of value whilst means seek to fulfill them, we need to recognise that our ends are constantly affected by our means, and that some ends are more consistent than others. I thus suggest that rather

than taking means-end rationality itself to be a value that is undermined by cognitive error, we should take the consistency of ends to offer incremental degrees of value in the normative judgment of cognitive error. This perspective is built into the positive thesis I will offer below.

Let me summarize the point I have reached so far, then, in critically considering different accounts of cognitive bias and fallacy. Cognitive errors cannot all be understood as logical errors, and those that can be thus understood can also alternatively be interpreted as mistaken assumptions. These assumptions cannot be usefully classified as 'false', because mistaken assumptions about what is true do not necessarily involve falsity. If we understand them as breaches of social normativity, it raises a problem of relativism, but the analysis of fallacies from the point of view of their social function usefully points us to a dialectical function for argument that is undermined by fallacy. Cognitive errors can also not be consistently understood in terms of means-end rationality where the ends remain stable, but they can be understood more broadly as disruptions to the consistency of ends.

What the last three approaches all suggest is that cognitive biases and fallacies consist of assumptions. The question is then what sort of assumptions, and how these can be generally characterized. The idea that both cognitive biases and fallacies are disruptions to a dialectical process, and that this dialectical process is one that can make our ends more consistent, is one that will now be incorporated into my wider proposals for a view of cognitive biases and fallacies as absolutisations.

Cognitive biases and fallacies as absolutisations

My positive thesis about the shared properties of cognitive biases and fallacies is that *both fallacies and the normative element of cognitive biases consist in absolutisations*.

By an *absolutisation*, I mean the assumption of a claim, explicitly or implicitly, that is taken to represent a completely true or false state of affairs. Absolutised beliefs lack *provisionality* and *incrementality*. By provisionality, I mean a psychological state in which alternatives to the belief concerned can be considered, which requires alternatives to be meaningful and to be potential objects of attention. By incrementality, I mean the possibility of conceptualizing our claims as part of a spectrum, whether that is a spectrum of variable qualities or one of probability.

Verbal tests of absolutisation may go a long way towards identifying it, though they are an approximate tool given that it is a psychological state that can be expressed in varying ways. A verbal test for absolutisation might involve asking whether or not a claim that we believe in can be interpreted incrementally: if it cannot, it is likely to be absolute. For example, 'God exists', 'the book does not exist', and 'the dog is black' are all apparently absolute claims. All of them might potentially be understood or interpreted in incremental terms, but when a person subject to a cognitive bias or fallacy *cannot* express them in incremental terms, they are liable to be absolutised. On the other hand 'I strongly experience something that I call God', 'I haven't managed to find the book so far' and 'the

colour of the dog's coat tends towards the black end of the gray scale' are all more explicitly incremental statements. The absolute statements in each of these cases are either true or false, with no possibility of degrees of justification. Even a dog with one gray hair makes 'the dog is black' false. If, however, we understand these claims in incremental ways that focus on the degree of justification we can gain from experience so far, and that implicitly acknowledge the possibility of a degree of error, they are far more likely to be held provisionally.

To show that absolutisation is a feature of every fallacy and cognitive bias requires a survey of every fallacy and cognitive bias – which is obviously beyond the scope of this paper. However, I have performed this survey elsewhere (Ellis 2015), where I have identified 189 fallacies and cognitive biases, all of which are amenable to an analysis in terms of absolutisation. Here I will provide a selection of examples to merely illustrate the ways in which absolutisation can account for a range of cognitive biases and fallacies.

Examples of absolutisation in cognitive biases

My first example is perhaps the most obvious and basic cognitive bias: confirmation bias. Rather than surveying evidence and drawing the best available conclusions that fit that evidence, our tendency is to pre-judge conclusions that fit our purposes and then find or construct reasons to support them, often taken very selectively from a context of ambiguous evidence that could potentially support a number of interpretations. For example, when asked to identify the rules behind number sequences, people overwhelmingly seek to confirm rather than falsify their beliefs about the rules (Wason 1960).

Confirmation bias is absolutising because, whilst we are subject to it, there is no possibility of modifying the conclusion we are seeking to prove in the light of evidence that does not confirm it – rather we ignore any such potential evidence. Thus, even if the conclusion is expressed in terms of incremental, even measurable, qualities (e.g. there are 2 liters of water in this bucket) it is possible to maintain it absolutely merely by excluding possible counter-evidence (e.g. the possibility that the bucket manufacturer marked the scale on the bucket inaccurately). Much of the time we have much practical justification for ignoring possible counter-evidence, but it obviously becomes more salient when it is of practical or epistemic importance, or when counter-evidence becomes at least a potential object for our attention. Here, then, we can notionally distinguish a descriptive element of confirmation bias that is merely an effect of our embodied limitations, from a normative element that might be corrected. Normally it is neither relevant nor important to query the accuracy of the scale on a bucket or any other measuring device, but there may be some occasions when such questioning becomes much more relevant and important, and normative responsibility for recognising the challenges to the scale requires us to recognise and overcome confirmation bias.

Attribute substitution is another type of cognitive bias that can readily be interpreted in terms of absolutisation. This consists in the substitution of an easier question (or easier explanation) in substitution for a harder one. For example, students first questioned about the number of their recent dates and then about

their level of happiness, tended to think about the level of happiness in terms of the number of dates rather than engaging with other (more complex) aspects of their happiness (Kahneman 2011, 97-104). Sunstein (2005) has also found these attribute substitutions used in moral judgments, in the form of appeals to moral authority or reliance on stereotypes substituting for engagement with moral complexity. Absolutisations are conceptual shortcuts that make judgment much quicker and easier than engagement with the complexity created by incrementality, so the absolutisation of happiness and of moral justification accounts readily for the normative element of the cognitive bias in each of these studies. Sometimes circumstances demand that we make such shortcuts, but at other times we have more potential flexibility.

There is a cluster of cognitive biases associated with assumptions about agency or its absence (the taking or avoiding of responsibility), one example of which is the illusion of control. Subjects tend to over-estimate the causal relationship between their own action (e.g. pressing a button) and a correlated event (e.g. a light coming on), even when they have been warned that there might not be any causal relationship (Allan & Jenkins 1980). In this case, it is our control and choice that are absolutised, as we find it much harder to engage with the idea of intermittent, partial, or unreliable responsibility for the effects of our actions. To some extent such over-estimation of our causal input may be an inextricable part of our embodied functioning, but there is nevertheless a normative component. That such a normative component can shift is demonstrated whenever somebody faces up to previously ignored moral responsibility for the effects of one of their actions.

Another example of a cognitive bias, which I will select here precisely because it may initially seem inhospitable to my thesis, is anchoring. Anchoring has already been mentioned and referenced above, and consists in our tendency to be primed by immediate past experience into giving attention to only a limited range of options framed by that immediate past experience. This is an effect that salespeople can use to their advantage by starting negotiations at a high price that then serves as the 'anchor', so that the purchaser feels that they have driven a bargain when they obtain reductions on it, even though the price is still unduly high in terms of the wider market.

To a large extent this cognitive bias appears to be merely a description of an unavoidable effect. One could understand such an effect in terms of the activation of neural connections by an anchor, together with our neural limitation of being unable to move immediately to a completely different set of activated connections. Rather we can only modify and add to the connections already activated, at least until this activation has died down. However, such a description of the power that anchoring has over us does not preclude a normative element, and some research has found ways that people can adjust and compensate for the limitations of anchoring. Simmons et al (2010) found that when they were aware of the direction in which they needed to adjust their anchor, people were able to effortfully compensate for the limitations of an anchor. We are indeed limited by a need to modify our existing neural activations, but that does not necessarily doom us to only modifying them in one particular way or to a particular limited extent.

The absolutisation involved in anchoring, then, is relevant only to the normative element of that bias (however large or small that normative element might turn out to be). We could absolutise the limitations that anchoring places on us by thinking of them deterministically, and in that way limit our potential incremental adjustment responses. At the other extreme we might imagine ourselves entirely free of anchoring, and able to create entirely new anchor points regardless of the previous priming of our neural connections, and this would involve the opposite error of absolutising an adjustment response that can only be incremental. Absolutisation does not account for the whole of the bias as psychologists have described it, but it does account for the adequacy or inadequacy of our practical response to it.

Examples of absolutisation in fallacies

Absolutisations can account for fallacies in the same way, with the difference only that, fallacies being entirely normative, absolutisation accounts more fully for the whole phenomenon. As I have already argued above, fallacies can be understood as assumptions that interfere with the dialectical function of argument and with the consistency of ends, and my argument here is that absolutisation is the mechanism that interferes in these respects.

Let me first take the example of a formal fallacy such as affirming the consequent. Here, as I have already argued above, the logical error can be alternatively understood as the adoption of an assumption, which in the case of affirming the consequent would take the form *'If P then Q' also entails 'if Q then P'*. In a logical error, the problem involved in adopting such an assumption is not that it is necessarily a false assumption, but that it remains unacknowledged, and that the conclusion is claimed to follow merely from the stated premises when this further hidden premise is also required. I want to argue here that it is this lack of acknowledgement that absolutises the assumption. In psychological terms, it is a *repressed premise*. A repressed content is absolutised because we are unable to apply the conscious processes of incremental assessment to it. It then creates a disruptive influence, leading us to inappropriate conclusions and presenting those conclusions, not as merely possible, but as inevitable.

If we were alternatively to apply Walton's account of fallacies as disrupting a dialectical process in which those with opposed views seek to reach agreement by appealing to each others' commitments, we could also see even formal fallacies as operating in this way. Logical reasoning is socially and psychologically important because it allows prior commitments to be consistently applied. However, a repressed premise interferes with that process and creates or perpetuates inconsistent commitments, insulating those inconsistent commitments through absolutisation.

Let us take the example of affirming the consequent and put it in a practical context. First, a reminder of the argument with its fallacious assumption:

- If P then Q
- ('If P then Q' also entails 'if Q then P')
- Q
- Therefore P

Now, suppose Harry and Harriet are debating the ethics of fox-hunting. Harry makes an argument that follows the pattern of affirming the consequent: "Fox-hunting is embedded in the culture of rural England" he says. "I know you live in the village, Harriet, but if you were a proper country person you would be in favor of fox-hunting." Analyzed, this argument would take the form:

- If you are in favor of fox-hunting then you must be a country dweller.
- (assumed: the above must also imply that those who are country dwellers must be in favor of fox-hunting)
- Harriet is a country-dweller.
- Therefore Harriet must be in favor of fox-hunting.

Harriet, however, replies: "That's stupid and prejudiced. Just because you live in the countryside doesn't mean you have to approve of torturing innocent animals". The debate makes no further progress, because Harry is insisting on his fallacious assumption, and the two can thus find no common commitments on the basis of which to reach any agreement. The debate is blocked because of the absolute nature of Harry's assumption. If he only recognized that the association between country-dwelling and support for fox-hunting was an incremental rather than an absolute one, there might be a basis for further discussion in which the pros and cons of fox-hunting might actually be examined.

In addition to the formal fallacy of affirming the consequent, the above example could also be interpreted in terms of various informal fallacies. For example, a causal rather than logical interpretation of Harry's claims about the relationship between fox-hunting and country-dwelling might reveal a fallacy of mono-causal explanation. Again, this involves the absolutisation of one cause and a lack of acknowledgement of the incremental contribution of more than one cause to a given outcome. It might also be interpreted as an *ad hominem* fallacy, in which Harry makes the mistake of assuming that Harriet's personal characteristic of being a country dweller is necessarily relevant to the justification of her beliefs about fox-hunting, when there is at best a highly contingent relationship. Again, there, it is not the assumption of the possible relevance of personal characteristics of the arguer to an argument that disrupts any dialectical process here, but rather the absolute nature of the assumption being made.

The common pattern of absolutisation in both formal and informal fallacies involves the treatment of inductive reasoning as though it was deductively valid. Whilst inductive reasoning involves a pattern of generalization from finite evidence, thus offering only incremental degrees of justification, deductive reasoning appears to offer certainty of the linkage between premises and conclusion (even though the premises themselves lack certainty). If it is true that Socrates is a man and men are mortal, we are traditionally told, then we can be certain that Socrates is mortal. When this apparent certainty is wrongly attributed to conclusions that are merely a matter of incremental justification, they are absolutised.

Absolutisation is a more adequate way of characterizing the problem with fallacies, not only because it shows their common ground with cognitive biases, but also because it offers us a more effective basis of judgment in the many doubtful cases we encounter beyond the constructed clarity of textbook examples. For example, to distinguish between an ad hominem argument and a jocular remark with a personal edge, one needs to consider not whether a conclusion is necessarily supported by its premises, but rather whether the social process of finding common practical grounds of action has been obstructed by what has been said. The interpretation of both parties in relation to social expectations and emotional states is at stake here, and determinate judgments could not be reached about specific examples without this contextual information. However, absolutisation can be used as a criterion by considering whether the person offering an argument held absolute beliefs about its conclusion. A jokey remark is recognized as limited and contextual, within a field of play, whilst an ad hominem argument involves absolute beliefs about a person that are then used to dismiss their beliefs. Of course, a remark that was intended by one party as jocular and non-absolute may be interpreted by its recipient (or by a third party) as ad hominem, but in this case the absolutisation (and the responsibility for it) lies with the interpreter.

Cognitive errors, dogmas and metaphysical beliefs

So far, then, I have argued that fallacies of all kinds consist in absolutisations, and that the normative element of cognitive biases also consists in absolutisation. Absolutisation may often be identified philosophically, through consideration of argument, but consists in the psychological state of an individual when they offer a claim (implicitly or explicitly) without awareness of incrementality in its justification and thus without awareness of the possible incremental justification of opposing claims. Since psychological states are constantly changing, we need to consider a freeze-frame version of the state at the time a claim is made.

Such a characterization of fallacies and cognitive biases thus further requires a characterization of dogmatism as a psychological state. Absolutisations are produced by people whose psychological state is, at the time of their claim, dogmatic. In a dogmatic state, other alternatives are not available to us because we have absolutised one claim and repressed alternatives. However, that is likely to be a temporary characterization of a psychological state. At another time, the repressed alternatives may emerge and be embraced. Provisionality, as an alternative state, involves the simultaneous recognition of alternative beliefs with varying degrees of justification. Incremental models are crucial for provisionality, because they allow degrees of justification to be compared, but they are incompatible with dogmatic states because alternatives cannot be repressed whilst their justification is recognized as partial.

A dogma, then, is a belief that we hold with the repression of alternative options. The presence of a dogma cannot be known determinately from the words used to express it, but where the terms used to express a belief are absolute, it is very likely that the associated judgments and psychological states will also be absolute. Awareness is needed of the possibility of non-absolute interpretations in practice even when a person's language seems very clearly absolutist.

Bearing in mind this caveat, however, we nevertheless need a term for the kinds of statements that instantiate and typify dogmatic belief. Such a term is already present in philosophy, and largely lines up with the psychological function I am discussing, even if some qualifications are needed in using it: that term is metaphysics. When used (as it often is) in a sense equivalent to ontology, metaphysics consists in claims about ultimate existence or non-existence. Such claims are not open to incremental enquiry, in the sense that judgments made about them must repress alternatives, rather than hold alternatives in mind as incrementally justified along with the assertion made. It needs to be noted that 'metaphysics' here is not equivalent to 'prior or categorial assumption'. Such assumptions are only dogmatic if they exclude the possibility of alternative possibilities. For example, to assert Kantian categories as prior conditions for experience is only dogmatic if we insist that the Kantian transcendental deduction is the only possible one.

Given this stipulation as to the nature of metaphysics, however, it then becomes possible to draw out the important implications of cognitive error as absolutisation for philosophy. Whilst I appreciate that my argument so far only establishes that cognitive error can usefully be understood as absolutisation, my thesis involves the wider claim that all absolutisations are cognitive errors. This is not a thesis that can be established by a priori argument, but needs to be understood as a theory that can be used not only to explain existing evidence about the common factors in cognitive bias and fallacy, but also make fruitful predictions about cognitive errors of all kinds that interfere with our judgments.

If all absolutisations are cognitive errors, then we should include not only fallacies and the normative element of cognitive biases as cognitive errors, but also metaphysical beliefs. Amongst metaphysical beliefs we should count not only the most sweeping positive metaphysical assertions in philosophy ('God exists', 'there are ultimate and independent material entities', 'minds exist distinct from bodies' etc) but also their negations ('God does not exist'), unconditional statements about ordinary objects ('The book exists'), and claims involving absolute and impermeable boundaries ('Italians are joyful, but Finns are gloomy'). In all such examples, as already stated, we must allow for the possibility of the language being interpreted non-absolutely, but when interpreted absolutely (as it commonly is, especially when discussed philosophically), these claims involve absolute assumptions. They also cannot be held provisionally, because their acceptance, in their own terms, necessarily involves the exclusion of alternatives.

For example, let us take the example 'God exists'. This belief cannot be held provisionally as long as 'God' is interpreted to mean an independently existent, infinite being (as opposed to, say, the finite and dependent object of religious experience). Such existence cannot be a matter of degree – God cannot partially exist – and thus is either accepted as justified or rejected. In accepting such a claim we have to repress any justification that can be given for the converse metaphysical belief, 'God does not exist'.

Given that fallacies and the normative element of cognitive biases are processes of reasoning that involve absolutising assumptions, the recognition that these

absolutising assumptions are metaphysical can then be fruitfully fed back into the analysis of fallacies and cognitive biases. I have found metaphysical assumptions in every example of fallacy or cognitive bias that I have examined, and elsewhere use metaphysical assumptions as the basis of a comprehensive categorization of cognitive errors (Ellis 2015).

Practical responses to cognitive biases and fallacies

Finally, then, I want to argue that this understanding of cognitive biases and fallacies as absolutisations provides us with the basis for a much more effective practical response to them, whether we envisage that practical response as occurring within in an individual practice, in a therapeutic setting that aims to correct the damaging effects of cognitive biases in individual lives, or in the social settings of organizations that seek to improve judgment and decision-making.

For wider understanding and engagement with cognitive biases, they need to be reducible to a few relatively simple common elements. The bewildering complexity and heterogeneity of cognitive biases as they are presented at present does not aid wider engagement with them, whilst the success of Kahneman (2011) in providing even a partial explanatory framework that covers a range of cognitive biases provides an indication of the need for wider and more comprehensive explanatory frameworks. 'Fast thinking' is a helpful model for what we need to avoid up to a point, but given the ubiquity and necessity of such fast thinking in a range of situations, it does not sufficiently isolate the type of thinking we need to avoid. 'Absolutising thinking' would be a more adequate way of characterizing the type of thinking we need to avoid, because it focuses on the reason why fast thinking is sometimes inappropriate – namely because it represses awareness of alternatives. Those alternatives may or may not be the ones we should adopt in the circumstances, but it is our lack of awareness of them as possibilities that we might have considered that limits the effectiveness of our responses.

I thus propose absolutisation as the key normative feature of cognitive biases that needs to be adopted in further presentation, research, and discussion of them. It has the advantages of simplicity, scalability, adequacy, and breadth. It is simple in offering one core feature for all cognitive biases and fallacies – in effect all of them can be reduced to one cognitive error. It is scalable, because that one cognitive error can also be expanded into more detailed accounts of every type of cognitive error, whilst providing the basis of the categorization scheme suggested above to bring them together. It is adequate because it focuses on the normative elements of fallacies and cognitive biases that need to be central to our practical response to them. Finally, it has the advantage of a breadth that breaks down unnecessary disciplinary boundaries between philosophy and psychology, allowing both philosophers and psychologists to continue to research this area with a common understanding of its significance and application.

Once it is agreed that cognitive biases and fallacies should be primarily understood in this way, the remedies become clearly those that address absolutisation. Broadly I suggest that this can be done at three levels. Firstly at the level of explicit belief or socially shared beliefs, one can develop critical

thinking responses in which we are more effectively trained in recognising the limitations of absolutising assumptions. Procedures adopted by organizations can incorporate the need for reflection on absolutising assumptions, whilst academic training could incorporate explicit critical thinking training that questions absolutising assumptions to a much greater extent. Secondly, at the level of meaning, we can engage in activities that limit our tendency to absolutise particular cognitive models and make absolutised representational assumptions about those models: the arts would provide the primary field of such practice, encouraging awareness of the provisional and ambiguous nature of symbols. Thirdly, we can work at the level of individual awareness, adopting practices that encourage wider immediate awareness of alternative models and reduce the likelihood of us absolutising a particular belief by training us into habits of broader awareness. Meditation and mindfulness practices provide the prime example of this type of practical response. All these types of practical response have been used in therapeutic settings, but have largely been ignored in wider philosophical and scientific presentation of cognitive biases and fallacies. Kahneman's remarks quoted at the beginning of the paper about the absence of a 'warning bell' to alert us to the presence of cognitive error just do not sufficiently engage with the range of practical options open to us, because of a failure to identify the core tendencies to be practically addressed.

Conclusion

My overall argument here began by rejecting both libertarian and deterministic assumptions to the understanding of cognitive biases and our responses to them. If we avoid these kinds of assumptions (which are absolute and metaphysical in nature, and thus involve cognitive errors), we are left with an experience of responsibility for cognitive error that should not be unnecessarily separated from our experience of moral responsibility. This requires us to identify clearly *what is wrong* with both the normative element of cognitive biases and with the fallacies identified in philosophical tradition.

I then argued that the traditional identifications of *what is wrong* with cognitive error – focusing on logical mistakes, falsity, social deviance or failure of means-ends rationality – are at best incomplete, none offering an entirely satisfactory explanation for this normativity. The partial explanations they do provide can instead be incorporated into an account of cognitive error as absolutisation, where this absolutisation is understood as blocking a dialectical process of either psychological or social integration and thus preventing both consistency of ends within individuals and the resolution of conflict at a social level. I then demonstrated the ways in which absolutisation can be found in the normative element of a range of example cognitive biases and fallacies.

The implications of absolutisation as the key to *what is wrong* with cognitive errors were then explored. If absolutisation is the problem, we can find that same problem in metaphysical dogmas that make these same absolutising assumptions explicit in the philosophical realm. The range of metaphysical beliefs can also then provide an organizing principle to help us understand the common factors in a range of cognitive errors. Understanding cognitive error as absolutisation thus provides us with new and more adequate ways forward in

coherently presenting cognitive errors and helping to prepare people in a range of situations in making adequate practical responses to them.

Bibliography

Allan, L., and Jenkins, H. (1980). The judgment of contingency and the nature of the response alternatives. *Canadian Journal of Psychology* 34(1), 1-11

Ellis, R. M. (2015). *Middle Way Philosophy 4: The Integration of Belief*. Raleigh: Lulu

Evans, J, Barston, J., and Pollard, P. (1983). On the conflict between logic and belief in syllogistic reasoning. *Memory and Cognition* 11(3), 295-306

Fujita, F., & Diener, E. (2005). Life Satisfaction Set Point: Stability and Change. *Journal of Personality and Social Psychology* 88(1), 158-164

Hauser, M. (2006). *Moral Minds: How Nature Designed our Universal Sense of Right and Wrong*. New York: Harpercollins

Jha, A., Krompinger, J. & Baime, M. (2007). Mindfulness training modifies subsystems of attention. *Cognitive, Affective and Behavioral Neuroscience* 7(2), 109-119

Johnson, M. (2014) *Morality for Humans: Ethical Understanding from the Perspective of Cognitive Science*. Chicago: University of Chicago Press

Kahneman, D. (2011). *Thinking, Fast and Slow*. London: Penguin

Kant, I., trans. Lewis White Beck (1785/1959). *Foundations of the Metaphysics of Morals*. Upper Saddle River: Prentice-Hall

Siegel, D. (2007). *The Mindful Brain*. New York: Norton

Simmons, J., Leboeuf, R. & Nelson, L. (2010). The Effect of Accuracy Motivation on Anchoring and Adjustment: Do people adjust from provided anchors? *Journal of Personality and Social Psychology* 99(6) 917-932

Sunstein, C. (2005). Moral Heuristics. *Behavioral and Brain Sciences* 28(4), 531-542

Taleb, N. (2008). *The Black Swan: The Impact of the Highly Improbable*: London: Penguin

Wason, P.C. (1960) On the failure to eliminate hypotheses in a conceptual task. *Quarterly Journal of Experimental Psychology* 12(3), 129-140

Walton, Douglas (1987) *Informal Fallacies: Towards a Theory of Argument Criticisms*. Amsterdam/Philadelphia: John Benjamins