

Merited Irrationality and the Functions of Belief

Allan Hazlett

1. THE PROBLEM OF THE SOURCE OF EPISTEMIC NORMATIVITY

There is a kind of evaluation of beliefs that we call “epistemic”: when we engage in this kind of evaluation, true beliefs are better than false beliefs, rational beliefs are better than irrational beliefs, and beliefs that amount to knowledge are better than beliefs that don’t. Epistemic evaluation is normative: in epistemic evaluation we do not merely classify a person’s beliefs, we arrange those beliefs in a hierarchy: in epistemic evaluation, to say of a belief that it is true is to give it a positive evaluation; to say of a belief that it is false is to give it a negative evaluation. Given the fact that epistemic evaluation is normative, there arises a question about the “source” of this normativity. Here’s what I mean.

We just saw that epistemic evaluation involves, at least, the evaluation of beliefs vis-à-vis the following rule:

TRUTH RULE: Someone’s belief that p is better if p is true, and worse if p is false.

This is compatible, of course, with there being false beliefs that epistemically come out quite well, all told, and with there being true beliefs that epistemically come out quite bad, all told. And, furthermore, it is compatible both with the view that truth is the primary or fundamental epistemic good, and with the view that there are a plurality of primary or fundamental epistemic goods.

Now, compare the evaluation of beliefs vis-à-vis the TRUTH RULE with evaluation of beliefs vis-à-vis the following:

TUESDAY RULE: Someone’s belief that p is better if she formed that belief on a Tuesday, and worse otherwise.

In other words, compare epistemic evaluation with a hypothetical kind of evaluation featuring the TUESDAY RULE, where beliefs formed on Tuesdays are evaluated positively, and beliefs formed on other days of the week are evaluated negatively. We are inclined to say, at least at first glance, that evaluation of beliefs vis-à-vis the TUESDAY RULE is somehow inappropriate (or illegitimate, or unwarranted, or mistaken). And more: we want to say that evaluation of beliefs vis-à-vis other non-epistemic norms is inappropriate, in the same sense in which epistemic evaluation is appropriate. Consider the familiar contrast between epistemic evaluation of beliefs and prudential or pragmatic evaluation of beliefs, e.g. the evaluation of beliefs in terms of the psychological comfort enjoyed by the believer, as a result of her believing as she does.

The problem of the source of epistemic normativity (or the *source problem*, for short) is to answer the following two questions:

- (i) Is it the case that evaluation of beliefs vis-à-vis epistemic norms (like the TRUTH RULE) is appropriate (in some sense), and that evaluation of beliefs vis-à-vis non-epistemic norms (like the TUESDAY RULE) is inappropriate (in the same sense)?
- (ii) If so, in what sense?

It is helpful to compare the source problem with what epistemologists call the “value problem.” This problem involves answering the question: in what way (or how, or why, or in what sense) is knowledge more valuable than (mere) true belief? In the *Meno* Socrates presents this as a problem arising in the practical domain: one who has a true belief about the way to Larissa, not amounting to knowledge, seems to be in no worse a position, when it comes to getting to Larissa, than one who knows the way to Larissa. But this isn’t right: knowledge has a kind of stability and security, compared to true belief, and it is therefore of greater practical value. Contemporary epistemologists typically consider a variant of Socrates’ problem, but in the epistemic domain. The contemporary value problem concerns what additional epistemic value knowledge has, in addition to the epistemic value of true belief. It is this version of the question that, for example, gives rise to the “swamping problem” for externalist theories of justification (Kvanvig 2003, pp. 44-51); true belief must be assumed to have some epistemic value, if said value is going to “swamp” the value of knowledge.

I have two points here. First, the source problem is distinct from the value problem. This can be seen from the fact that the question posed in setting up the value problem – how is knowledge more epistemically valuable than true belief? – presupposes the epistemic value of true belief. But an account of the epistemic value of true belief is precisely what is sought in solving the source problem. That is the second point: standard contemporary approaches to the value problem presuppose a (particular) solution to the source problem. So the latter is, in that sense, prior to the former.

There are at least two ways to respond to the source problem. The answer I prefer is PLURALISM ABOUT DOXASTIC EVALUATION: epistemic evaluation of beliefs is no more appropriate than any other way of evaluating beliefs. But this is counterintuitive for most philosophers. Most of us want to say that epistemic evaluation of beliefs is uniquely appropriate, appropriate in some way that other ways of evaluating beliefs aren’t appropriate. One common way of articulating that idea is to say that epistemic evaluation of beliefs is the evaluation of beliefs, qua beliefs. But what does this mean?

First, note that it is not the case that all x can be “evaluated qua x .” Samples of barium cannot be “evaluated, qua samples of barium.” There is no such thing as “the evaluation of barium, qua barium.” Barium, we might say, does not have any essential standards of evaluation – it has no standards of evaluation that flow from its essential nature, that flow from it, qua barium. So one of the things we are saying, if we say that epistemic evaluation of beliefs is the evaluation of beliefs, qua beliefs, is that beliefs are unlike samples of barium, in this respect: they *do* have essential standards of evaluation.

Second, note that we can give either a de dicto or a de re reading of “the evaluation of x qua x .” On the de dicto reading, there are standards of evaluation that flow from the essential nature of some kind, of which x is a member. Consider this individual lawyer, Kyra. We can evaluate Kyra qua lawyer, in the de dicto sense, by evaluating her vis-à-vis various norms the

flow from the essential nature of the kind *lawyer*. It flows from the essential nature of the kind *lawyer* that a good lawyer shall have such-and-such qualities, and whether or not Kyra has those qualities determines how she does when evaluated, qua lawyer. Importantly, in so evaluating her we make no assumption about whether Kyra is or is not essentially a lawyer. The standards of evaluation we employ flow not from *Kyra's* essential nature, but from the essential nature of the *kind* that we (correctly) describe her as being a member of. On the other hand, on the de re reading of “the evaluation of x qua x ,” there are standards of evaluation that flow from the individual essential nature of x itself, de re, and not merely relative to some description of x . This will happen when (i) there are standards of evaluation that flow from the essential nature of some kind F , and (ii) x is essentially F , i.e. F determines x 's essential nature. In such cases “the evaluation of x qua x ” will be the evaluation of x as the kind of thing that it essentially is.¹

Given all this, we must distinguish between two views:

DE DICTO EPISTEMIC ESSENTIALISM ABOUT DOXASTIC EVALUATION (in the sequel DE DICTO ESSENTIALISM): Epistemic evaluation of beliefs is appropriate, and non-epistemic evaluation of beliefs inappropriate, because epistemic evaluation is the evaluation of beliefs, qua beliefs, i.e. the evaluation of beliefs relative to standards of evaluation that flow from the essential nature of the kind *belief*.

DE RE EPISTEMIC ESSENTIALISM ABOUT DOXASTIC EVALUATION (in the sequel DE RE ESSENTIALISM, and later just ESSENTIALISM): Epistemic evaluation of beliefs is appropriate, and non-epistemic evaluation of beliefs inappropriate, because epistemic evaluation is the evaluation of beliefs, qua beliefs. In other words, epistemic evaluation is the evaluation of beliefs relative to standards of evaluation that flow from the essential nature of the kind *belief*, and all beliefs are essentially beliefs. Thus epistemic evaluation of beliefs is the evaluation of beliefs as the kinds of things that they essentially are.²

I think the prospects for DE DICTO ESSENTIALISM are dim. Consider Kyra. In what way is evaluating Kyra vis-à-vis standards that flow from the essential nature of the kind *lawyer* appropriate, compared to all other ways of evaluating her? Kyra is, by hypothesis, a lawyer, so evaluating her vis-à-vis standards that flow from the essential nature of the kind *lawyer*

¹ It may also be possible that there are standards of evaluation that flow directly, as it were, from the de re essential nature of x , and not via any kind that x is essentially a member of. But this is not plausible in the case of beliefs – if there are standards of evaluation that flow from their de re essential nature, then this is because the kind *belief* has some essential standards attached to it.

² Note that DE RE ESSENTIALISM and DE DICTO ESSENTIALISM (which propose non-PLURALISTIC solutions to the source problem) are both committed to the following MINIMAL ESSENTIALIST THESIS (which says nothing explicit about the source problem): Epistemic evaluation is the evaluation of beliefs relative to standards of evaluation that flow from the essential nature of the kind *belief*.

seems better than, for example, evaluating her vis-à-vis standards that flow from the essential nature of the kind *plumber* (assuming Kyra isn't also a plumber). But Kyra may be (and if she is anything like most of us, will be) many things other than a lawyer – an athlete, a poet, a mother, a citizen, a lover, a cook, a friend, and so on. Each of these kinds may come with its own standards of evaluation. It would obviously be a mistake, when it comes to evaluating Kyra, to think that the standards that flow from the essential nature of the kind *lawyer* have any special status, relative to the various other standards that flow from the essential natures of the various other kinds that Kyra is a member of.³

The upshot of this is that lawyers may be much more than lawyers, and their property of being lawyers has no pride of place among their other properties. Beliefs, you may want to say, are not like that. But I see no plausible way of saying that without saying that beliefs are essentially beliefs. Thus, it seems that DE RE ESSENTIALISM is the best bet for those seeking to vindicate the idea that the epistemic evaluation of beliefs is uniquely appropriate, on the grounds that it is the evaluation of beliefs qua beliefs. In what follows, then, I'll confine my attention to versions of DE RE ESSENTIALISM (in the sequel, ESSENTIALISM).⁴

2. TWO SOLUTIONS SHELVED

In this paper I am concerned with a particular species of ESSENTIALIST solution to the source problem, which appeals to the idea that being true (or having some other positive epistemic status) is the *biological function* of belief.⁵ This solution should not be confused with two others, which I'll describe here as briefly as possible.

Some philosophers argue that the concept of belief is normative. (This is often, though it need not be, part of a broader argument that the concept of content is normative, or that all mental concepts are normative.) It is a necessary conceptual truth, so the argument goes, that being true is the (or at least a) “constitutive standard of correctness” for belief; alternatively truth is belief's “constitutive aim.” This explains why evaluation vis-à-vis the TRUTH RULE is appropriate, while evaluation vis-à-vis the TUESDAY RULE isn't. Call this view NORMATIVE ESSENTIALISM. On this view the normativity of belief is conceived of as irreducible (by contrast with another view, which I'll describe in a moment, and with the functionalism that is the topic of this paper; NORMATIVE ESSENTIALISM is neutral as to whether this poses a threat to metaphysical or methodological naturalism. It is defended by Daniel Dennett (1971), Donald Davidson (1983), Ralph Wedgwood (2002), Paul Boghossian (2003, 2005), Nishi Shah (2003), and Shah and David Velleman (2005).

³ You might want to say that the appropriateness of evaluating someone vis-à-vis some standard depends on what the person you are evaluating thinks: what she desires, intends, or cares about. See PERFORMANCE ESSENTIALISM, below.

⁴ ESSENTIALISM is formulated in terms of epistemic and non-epistemic evaluation. Above we gave a vague and informal characterization of “epistemic” evaluation, by giving a list of evaluative practices (true belief better than false, rational belief better than irrational, knowledge better than non-knowledgeable belief). Perhaps a more rigorous account of the “epistemic” is needed. Note that this problem, if it is a problem, does not face PLURALISM.

⁵ In what follows I consider only the view that being true is the biological function of belief. Everything I say should apply, mutatis mutandis, to variants on which some other positive epistemic status is substituted for truth.

Some philosophers conceive of knowledge as an achievement, and as such something worthy of praise. So Ernest Sosa (2007, 2009, forthcoming) argues that knowledge is cognitive success attributable to the knower's epistemic virtue; John Greco (2003) argues that knowledge is true belief for which the believer deserves credit; Wayne Riggs (2002, 2009) argues that knowledge is an credit-worthy achievement. This suggests a solution to the source problem: reduce epistemic normativity to the normativity of intentional action or, as Sosa puts it, "performance." In general, when I set out to do something, or try to do something, or intend to do something, and don't do that thing, I deserve (at least other things being equal) a negative evaluation. I have *failed*, and failure is (in general, other things being equal) *bad*. Ditto, mutatis mutandis, if I *succeed*. To believe that *p*, on this view, is to attempt or endeavor to believe the truth; in this sense, truth is the (or an) "aim of belief." Hence the appropriateness of evaluation vis-à-vis the TRUTH RULE. Call this PERFORMANCE ESSENTIALISM. This view appears compatible with metaphysical and methodological naturalism, assuming the naturalistic credentials of the notions of trying, intending, etc.

In the context of the source problem, it will not do to say that some beliefs, but not others, aim at truth. For epistemic evaluation is intuitively appropriate for all beliefs. And thus the feature of beliefs that explains the appropriateness of epistemic evaluation must be present in all beliefs. Therefore, if we wish to vindicate the intuition that epistemic evaluation is uniquely appropriate for all beliefs, then we must say that the fact that truth is the (or a) aim of belief is a necessary truth about belief. (Cf. David 2001, p. 156, Kelly 2003, p. 625, and Grimm 2008, p. 742.) This jives with the idea that that epistemic evaluation is the evaluation of beliefs, qua beliefs, which gave us the name "essentialism" (§1).

The solution to the source problem that I'm concerned with in this paper – FUNCTIONAL ESSENTIALISM – is also a species of ESSENTIALISM. A key difference between NORMATIVE and PERFORMANCE ESSENTIALISM, on the one hand, and FUNCTIONAL ESSENTIALISM, on the other, is that NORMATIVE and PERFORMANCE ESSENTIALISM, if true, are knowable a priori. If it's a necessary conceptual truth that belief has a constitutive standard of correctness, this could be known a priori. And we don't find out that knowledge is a credit-worthy achievement by empirical investigation; we find this out by conceptual analysis. FUNCTIONAL ESSENTIALISM, by contrast, is intended as an empirical hypothesis about belief, knowable only a posteriori.

3. FUNCTIONAL ESSENTIALISM

David Velleman argues that "belief aims at truth," or, more precisely, that "to believe a proposition is to accept it with the aim of thereby accepting a truth." (2000, p. 251) This conception of the aim of belief "allows but does not require the aim of belief to be an aim on the part of the believer." (Ibid. p. 252) How could it be the case that I accept *p* "with the aim of thereby accepting a truth," without that being my aim? Velleman writes that:

A person can ... aim cognitions at the truth without necessarily framing intentions about them. Suppose that one part of the person – call it a cognitive system – regulates some of his cognitions in ways designed to ensure that they are true, by forming, revising, and extinguishing them in response to evidence and argument. (Ibid. p. 253)

Can we construct a solution to the source problem from this idea? Consider: beliefs are designed to be true, and thus evaluation of beliefs vis-à-vis the TRUTH RULE is appropriate. Beliefs aren't designed to be formed on Tuesdays, and so evaluation of beliefs vis-à-vis the TUESDAY RULE is inappropriate. Compare snow tires, which are designed to enable driving in the snow. It's appropriate to evaluate a set of snow tires in terms of whether they enable one to drive in the snow, because that's what they were designed to do. It's inappropriate to evaluate snow tires in terms of whether they enable one to drive in sand, because that's not what they were designed to do.

If beliefs are literally designed to be true, who designed them? Design requires a designer; the reason snow tires are "designed to enable driving in the snow" is that the people who manufactured them had driving in the snow in mind, as their goal or intention, when they manufactured the tires (among various other things they might have had in mind). Clearly, we should not say that the believer designs her beliefs to be true. For while there may be some sense in which this is true of some of the beliefs of some believers – as when, for example, I consciously deliberate about the question of whether to believe p and eventually come to the conclusion that p is true – it is definitely not essential to believing that the believer have designed her belief with the goal or intention of its being true, whatever exactly "design" might mean here. The same point applies to the idea that, say, the parents of believers necessarily "design" their offspring to have true beliefs.

The only plausible thing to say, if beliefs are literally designed to be true, is that God designed our beliefs to be true. Thus, only theists can embrace the proposal that beliefs are literally designed to be true.⁶ Metaphysical and methodological naturalists cannot go along with this idea. They are apt to prefer a non-literal understanding of "design." So how is the claim that beliefs are designed to be true to be understood, if "design" is not to be taken literally?

The next move would seem to be an appeal to the familiar, perhaps all-too-familiar, idea that something can be "designed" by natural selection. (This is what Velleman appeals to immediately following the passage quoted above.) Just as snow tires have their function (enabling one to drive in the snow) in virtue of the goals or intentions of their designer, certain biological entities have their function in virtue of ... well, something about their evolutionary history, we want to say. But we will need to say much more for this non-literal conception of design to be philosophically illuminating.

The reason we must say more here is that we run the risk of tacitly committing ourselves to a fundamentally mistaken interpretation of the theory of evolution by natural selection. There is a popular conception of said theory which is little more than traditional theism, but with amoral and impersonal "nature" doing the work previously done by God. On this popular misinterpretation, nature has purposes, and these determine the proper functions of organisms and their parts, as well as of ecosystems and other supra-organismal entities. Nature designs things to be a certain way, sometimes things go according to design (this is good), and other times not (this is bad). But all this talk of "nature's designs" and "nature's

⁶ Thanks to Daniel Howard-Snyder for pressing me on the importance of this point. Cf., obviously, Plantinga 2002.

purposes” is completely worthless to us, unless it can be rigorously reformulated in a way consistent with what evolutionary biology actually says.

Consider an intuitive sketch of the idea that being true is the biological function of belief. Quine famously expressed the premise behind the functionalist view by saying that creatures with unreliable faculties “have a pathetic but praiseworthy tendency to die before reproducing their kind.” (1969, p. 126) Dennett writes that “the capacity to believe would have no survival value unless it were a capacity to believe truths.” (1971, p. 101) And Velleman says, of beliefs, that “[t]heir guiding the subject when true is what confers advantages on him, and so it appears to be what beliefs were selected for, in the course of evolution.” (2000, p. 253n) Reliable belief formation confers evolutionary advantage on the believer. Thus, the biological function of belief is to be true.

An adequate account would say more. Ruth Garrett Millikan (1984) says quite a bit more, proposing a general theory of “proper functions” (pp. 17-50). Consider the barbels on a catfish, which enable the catfish to detect the presence of food. We want to say that the biological function of these barbels is to enable the catfish to detect food. Two important points are implied by Millikan’s account. First, whether F is the biological function of x is not a matter of whether x ’s performing F confers evolutionary advantage on the token organism of which x is a part. Rather, it is a matter of whether x ’s performing F involves doing that which historically, as a matter of fact, conferred evolutionary advantage on the ancestors of the token organism of which x is a part. Therefore, to say that the biological function of a catfish’s barbels is to enable the catfish to detect food requires, at a minimum, that the catfish’s ancestors enjoyed an evolutionary advantage on account of having been enabled to detect food with their barbels.

Second, F is the biological function of x only if the very existence of x is explained by the fact that doing F conferred evolutionary advantage on the ancestors of the organism of which x is a part. Thus an additional requirement on saying that the biological function of a catfish’s barbels is to enable the catfish to detect food is that it must be the case that the present existence of said barbels is explained by the fact that the catfish’s ancestors enjoyed an evolutionary advantage on account of having been enabled to detect food with their barbels.

Consider Millikan’s example of the proper function of a human heart. Pumping blood is the proper function of my heart, in virtue of the following three facts:

- (1) I have ancestors who had hearts that pumped blood,
- (2) that said hearts pumped blood contributed to the fitness of said ancestors (relative to salient competitors), and
- (3) that I have a heart is (at least partially) explained by (2).⁷

⁷ Facts (1) and (2) correspond to the first Millikanian point, above, and fact (3) corresponds to the second. These three claims correspond roughly to the three necessary and sufficient conditions proposed in Millikan’s theory of “direct proper functions” (1984, p. 28). The account I’m sketching is an extreme simplification of Millikan’s theory, but (I hope) it gets the spirit right.

In this case, the explanation is plausibly genetic: because having a blood-pumping heart contributed to the relative fitness of my ancestors, genes coding for blood-pumping hearts proliferated in my ancestors, resulting in the proliferation of blood-pumping hearts in my ancestors, resulting eventually in the fact that I have a heart. Given all this, if my heart fails to pump blood, then it fails to perform its proper function: it is diseased or deformed or otherwise malfunctioning. More generally, it may be the case that F is the proper function of x , even if x fails to perform F . Indeed, this may be the case even if x is unable to perform F . Thus for biological functions, ‘ought’ does not imply ‘can’. Furthermore, it may be the case that that F is the proper function of x even if most actual x ’s fail to perform F : the function of a sperm cell is to fertilize an egg, even though very few do so. (Ibid. p. 29) Thus it need not be statistically normal for members of a certain kind of thing to perform their biological function. Finally, as mentioned above, x ’s performing its biological function may not actually confer evolutionary advantage on the organism of which x is a part: my teeth may function properly even when I’m chewing up a poisoned steak.

I think it’s clear that, at least if we seek an understanding of the notion of a “biological function,” we must say more about what sort of explanation is acceptable, when we are explaining the existence of x by appeal to the fact that doing F conferred evolutionary advantage on the ancestors of the organism of which x is a part. Only a “biological” explanation will do. In the case of my heart, the explanation is genetic: the reason I have a heart is that I have genes that code for a heart, and the reason I have those genes is because it was advantageous for my ancestors to have genes that coded for a blood-pumping heart, because it was advantageous for them to have blood-pumping hearts. So, for this explanation to work there must exist genes that code for a blood-pumping heart. And, in general, a genetic explanation, showing that F is a function of x , will work only if whether or not x is F is determined by the genotype of the organism of which x is a part.

It is not clear whether we should admit any other sorts of explanations as “biological,” in the relevant sense. Songbirds, when all goes well, since a song typical for their species, but the song an individual bird sings is not determined by that bird’s genotype, but rather by the songs that that individual bird was exposed to as a juvenile. Singing a species-appropriate song thus seems, in some obvious sense, to perform a “biological” function: that of attracting a mating partner. (Millikan’s conditions for a proper function are met: present-day sparrow singing is explained by the evolutionary advantages gained by their ancestors singing the sparrow-specific song.) But whether a bird sings a species-appropriate song is a learned behavior, not a matter of genetics.

However, it remains unclear how to draw the line between “biological” functions and others: certainly *part* of the reason that present-day Australians sing an anthem is that (some of) their ancestors sang “Advance Australia Fair” – but is it the biological function of Australians to sing “Advance Australia Fair”? Confining ourselves to genetic explanations removes some of the sloppiness that results from allowing non-genetic explanations to count as “biological.”

Although we spoke above of *the* function of the catfish’s barbels, and of *the* function of the human heart, it seems clear that something could have a plurality of biological functions. Consider the channel catfish’s swim bladder, which functions both to properly regulate the catfish’s bouancy (so that it sinks to the bottom) and to enable the catfish to hear (vibrations

are passed from the bladder to the inner ear via various parts of the skeleton). So while it will sometimes make sense to speak of *the* biological function of something or other, it will often be a mistake, and in that case we should speak of a certain function as *a* function of that thing.

So much for biological functions, in general. We sought an essentialist solution to the source problem, one that appeals to the idea that being true is the biological function of belief. We can now formulate such a view:

FUNCTIONAL ESSENTIALISM: Epistemic evaluation of beliefs is appropriate, and non-epistemic evaluation of beliefs inappropriate, because epistemic evaluation is the evaluation of beliefs vis-à-vis standards that flow from the biological function of belief. More precisely, the biological function of belief is to be true. Since beliefs are essentially beliefs, the epistemic evaluation of beliefs is the evaluation of beliefs as the kinds of things that they essentially are – namely, entities whose biological function is to be true.⁸

FUNCTIONAL ESSENTIALISM is thus committed to three things:

- I. An empirical claim: that the biological function of belief is to be true.
- II. A metaphysical claim: that biological functions determine the essences of the things that have them.
- III. A claim in intellectual metaethics: that because the biological function of belief is to be true, the evaluation of beliefs vis-à-vis epistemic norms is uniquely appropriate.

We saw, above, a sketch of a defense of claim (I), in the quotations from Quine, Dennett, and Velleman. With a Millikanian conception of biological functions in hand, we can now bolster that defense. (Cf. Millikan 1984, pp. 17-19 and pp. 93-4.) The idea, roughly, is that:

- (4) I have ancestors who had true beliefs,
- (5) that said beliefs were true contributed to the fitness of said ancestors (relative to salient competitors), and
- (6) that I have beliefs is (at least partially) explained by (5).

Is this plausible? Let us grant claims (4) and (5), for now, and consider whether (6) is true – and whether it is true in the right way, i.e. whether a “biological” explanation of the fact that I have beliefs can be given by appeal to the fitness-enhancing virtues of having true beliefs. Is whether someone has true beliefs determined by her genotype? Perhaps in this sense: there might be genes that code for certain kinds of cognitive faculties, faculties of a kind that are actually reliable, and thus faculties the use of which will (if all goes well) result in a human being’s having many true beliefs. If having true beliefs contributed to the fitness of

⁸ This formulation assumes – and I will not challenge here – that the appropriateness of all epistemic evaluation can be explained by appeal to the fact that being true is the biological function of belief. Alternatively, the functional essentialist could say that belief has a plurality of biological functions, but that these are all and only “epistemic functions”, e.g. that belief’s functions include being true, being rational, amounting to knowledge, etc.

my ancestors, then the aforementioned genes would have proliferated among my ancestors, and this might explain, eventually, why I have belief-forming faculties. It seems plausible that at least some of the ways that we form beliefs are genetically determined in this way, in virtue of the genetic determination of the sense organs that we have, the ways that said organs are connected to our brains, and so on. Alternatively, perhaps the case of belief formation is like that of the songbird, where forming beliefs is (in some sense) a learned or acquired practice, where (if all goes well) I learn reliable belief-forming practices from my conspecifics.

As far as I know, no defender of the claim that being true is the biological function of belief has ever attempted to provide the details of either of these sorts of account. No one has sought evidence, for example, that certain genes code for reliable belief-forming faculties, or that belief-formation is a learned practice. In other words, no one has attempted to give a “biological” explanation of the fact that we presently have beliefs, premised on the claim that having true beliefs enhanced the fitness of our ancestors. The defenders of FUNCTIONAL ESSENTIALISM have maintained that having true beliefs enhanced the fitness of our ancestors, but have not sought to connect this supposed fact up, in a “biological” way, with the fact that we are believers. Thus the case for FUNCTIONAL ESSENTIALISM is similar to the case often made for the claim that that xenophobia is biologically natural for humans, on the grounds that being xenophobic would have contributed to the fitness of our ancestors. No one has ever tried to show, for example, that whether an individual human is xenophobic is determined by her genotype. No one has sought to identify genes that code for xenophobia. Claims about biological functions will always be contingent on the empirical success of the relevant biological explanation.

However, notice that, even if *all this works*, claim (I) has not yet been established. For claim (I) says that being true is *the* biological function of belief. Being true is unique, among things that beliefs can be, in being the biological function of belief. As we saw above, a plurality of biological functions is possible, as in the case of the catfish’s swim bladder. So, for all we have said, belief may have a plurality of biological functions as well. On the other hand, in the absence of any reason to think otherwise, if we are confident that being true is a biological function of belief, perhaps it is reasonable to conclude that being true is the unique biological function of belief.

In any event, given claim (III), FUNCTIONAL ESSENTIALISM must say that being true is *the* biological function of belief.⁹ If being true is the biological function of belief, and given claim (II), evaluation of beliefs vis-à-vis the TRUTH RULE will be the evaluation of beliefs qua beliefs. Evaluation vis-à-vis the TUESDAY RULE, on the other hand, will not be the evaluation of beliefs qua beliefs, because it is not a biological function of belief to be formed on a Tuesday. But to establish the uniqueness of epistemic evaluation, the functional essentialist must say that evaluation of beliefs vis-à-vis non-‘epistemic’ norms, of any kind, is never the evaluation of beliefs qua beliefs. Being true must be the unique biological function of belief. For ESSENTIALISM, by contrast with PLURALISM, says that epistemic evaluation has a unique status when it comes to evaluating beliefs.

⁹ Or, as above, that belief’s functions are all and only “epistemic functions”; see footnote [8].

The Millikanian theory of biological functions, sketched above, provides the tools to defend claim (II). Here is Millikan:

Suppose that by some cosmic accident a collection of molecules formed in random motion were to coalesce to form your exact physical double. [T]hat being would have no ideas, no beliefs ... This because the evolutionary *history* of the being would be wrong. For only in virtue of one's evolutionary history do one's intentional mental states have proper functions ... That being would also have *no liver, no heart, no eyes, no brain*, etc. This, again, because the history of the being would be wrong. For the categories "heart," "liver," "eye," "brain," and also "idea," "belief" ... are proper function categories, defined in the end by reference to long-term and short-term evolutionary history[.] [B]eliefs ... are not such because of what they do or could do. They are such belief of what they are, given the context of their history, *supposed* to do and of how they are supposed to do it. [...] Beliefs ... are *essentially* things that can be true or false, correct or defective. (1984, p. 93)

What it is to be a heart – the essence of being a heart – is to have a certain proper function (to pump blood), in virtue of having a certain kind of evolutionary history, and what it is to be a belief – the essence of being a belief – is to have a certain proper function (to be true), in virtue of having certain kind of evolutionary history. Furthermore (here is the metaphysical claim), biological functions determine the essences of the things that have them, and belief is one of these.

Claim (I) is a necessary a posteriori truth, if it is a truth at all. This is obscured by the claim that belief is a "proper function category," which suggests an a priori conceptual truth. But that the heart pumps blood is not an a priori conceptual truth; so neither should we think, of the proposed fact that the function of belief is to be true, that this is an a priori conceptual truth (as it is on NORMATIVE ESSENTIALISM).

4. OPENING SALVO: TWO NATURALISTIC FALLACIES

FUNCTIONAL ESSENTIALISM seeks to explain the appropriateness of epistemic evaluation by appeal to an empirical claim about belief. Since epistemic evaluation is normative, and since it evaluates true beliefs as better than false beliefs, it seems that FUNCTIONAL ESSENTIALISM seeks to explain why true beliefs are better than false beliefs by appeal to an empirical claim about belief. And since that empirical claim is the claim that being true is the biological function of belief, it seems that FUNCTIONAL ESSENTIALISM seeks to explain why true beliefs are better than false beliefs by appeal to the fact that true beliefs are, in some sense, more natural than false beliefs. True beliefs are, to speak very metaphorically, what Nature intended. FUNCTIONAL ESSENTIALISM seems to conclude, on those grounds, that true beliefs are good.

There are two problems with this, if this is a correct characterization of what the FUNCTIONAL ESSENTIALIST is up to. The first concerns, in particular, concluding that something is good on the grounds that it is natural, and bad on the grounds that it is unnatural. The second concerns, in general, drawing normative or axiological conclusions from empirical descriptions of the world.

Consider the biological function of sexual intercourse. I hope it is relatively uncontroversial that among the biological functions of sexual intercourse – if this is not the unique biological function of sexual intercourse – is reproduction. This will certainly follow on the Millikanian conception of biological functions: the reason that sexual intercourse exists – the reason, e.g., that we have the organs that make it possible – is that having sexual intercourse proved evolutionarily advantageous to our ancestors, various sex-related genes were selected for, and so on. Tokens of sexual intercourse could thus be evaluated vis-à-vis standards implied by the biological function of sexual intercourse, namely, whether or not said token resulted in reproduction. Sexual intercourse is good, vis-à-vis this standard, when it results in reproduction, and bad otherwise. Call this *reproductive evaluation*.

So far, so good. But now imagine that someone maintains that this biological standard of sexual evaluation is the uniquely appropriate standard to apply when evaluating tokens of sexual intercourse, on the grounds that such evaluation is the evaluation of sex, qua sex. I suspect that (let me put this as mildly as I can) many of us would find that view counterintuitive. This for a variety of reasons.

The most basic way to put the point here is that there are a many ways of evaluating sex, other than reproductive evaluation and that these ways of evaluating sex are at least as appropriate as reproductive evaluation, if not very much more appropriate. So consider: evaluating tokens of sexual intercourse in terms of whether the intercourse was consensual (good) or not (bad); evaluating tokens of sexual intercourse in terms of whether the intercourse was pleasurable or satisfying to the parties involved (good) or not (bad); evaluating tokens of sexual intercourse in terms of whether the parties involved were married to each other (good) or not (bad); evaluating tokens of sexual intercourse in terms of whether one or more of the parties was committing adultery (bad) or not (good); and so on.

Now we have at least two options for diagnosing where exactly things go wrong, on the present proposal. We might make our argument metaphysical: the biological function of sex does not determine the essential nature of sex (not for creatures like us, anyway), and thus reproductive evaluation is not the evaluation of sex, qua sex, as claimed. (This seems at least partially right: someone who thinks that the unique purpose of sex is to reproduce has somehow missed the point; we want to say that such a person, in some sense, doesn't really know what sex is.) This opens up the possibility that the biological function of a kind might not determine its essential nature. In the sexual case, we can raise a challenge: why think of sex as essentially biological? Why not define sex in social or interpersonal or hedonistic terms?¹⁰

The other option would be to make our argument metaethical: to someone who criticizes non-reproductive sex as bad, qua sex, we could respond that we don't care only about whether the sex we have is good, qua sex. (We might not even care about this at all!) We might want to use sex, as it were, for something else. When hobos start a fire in a oil drum, they are unperturbed if their drum turns out to be bad, qua oil drum, because it lacks a crucial gasket. (This, too, seems a plausible response to the reproductive evaluation of sex: the importance of consent, in sexual intercourse, is not a matter of the essential nature of

¹⁰ See Nagel 1979 and Goldman 1977 for further discussion of the “nature of sex.”

sex, but it's even more important to (many of) us that tokens of sexual intercourse be consensual, than that they be good, qua sex (whatever that might end up meaning.) This opens up the possibility that evaluating x , qua x , is not always the uniquely appropriate way to evaluate x ; x might have other "functions."

The upshot of all this is that we should be leery of any attempt to deduce the goodness of true belief (and thus of any attempt to deduce the unique appropriateness of epistemic evaluation) from the fact that true belief is biologically natural, e.g. from the fact that the biological function of belief is to be true. This might be because belief is not properly understood as a biological kind (the metaphysical option, above) or because there are non-essential (e.g. non-biological) "functions" of belief (the metaethical option, above).

We sometimes use the word "function" in the following way: we say that F is the (or a) function of x whenever x 's performing F is something that want x to do, something that results in an outcome we desire. When we speak of "functions" in this way, the "function" of a thing is radically perspectival, relative to the interests of some individual or group. From the perspective of her firm, Kyra's function is to argue cases in court; from the perspective of her daughter, Kyra's function is to provide love and material wellbeing; from the perspective of the IRS, her function is to pay her taxes; and so on.¹¹ If "function" is used in this perspectival way, ESSENTIALISM is doomed. To ground the appropriateness of epistemic evaluation in the perspectivally relative "function" of belief would give us no reason to think that epistemic evaluation is *uniquely* appropriate: just as there is a perspective from which true belief appears best, so there is surely a perspective from which Tuesday-formed belief appears best.

That was the first prima facie problem with FUNCTIONAL ESSENTIALISM: deriving goodness from naturalness. The second prima facie problem is that the FUNCTIONAL ESSENTIALIST seems to be illegitimately deriving an 'ought' from an 'is': from a descriptive claim about the natural history of belief, she deduces a normative claim about the appropriateness of epistemic evaluation, which implies that true belief is better than false belief.

We must keep in mind the fact that whether F is the biological function of x is an entirely descriptive matter – at least if anything ever is an entirely descriptive matter. The biological function of a thing is a matter of its natural history. " F is the biological function of x " can be defined entirely in uncontroversially non-normative language. This is part of the appeal of such functional claims, for philosophical naturalists!

One way of seeing the problem here is by granting that being true is the biological function of belief, and then asking: why, therefore, is true belief better than false belief? Epistemic evaluation, as we said in §1, does not merely classify beliefs; it *evaluates* them positively or negatively, it says that some are *good* and others *bad*. Supposing Millikan et al. are right about the natural history of belief; how does this show us that epistemic evaluation is more appropriate than, say, "reverse epistemic" evaluation, which praises the false, the irrational, and the not-amounting-to-knowledge? That is the second prima facie problem with

¹¹ A similar relativity of functions arises within biology: from the perspective of a moray eel, the function of a cleaner shrimp is to remove parasites from the eel's mouth; from the perspective of an ecosystem, the function of nitrogen-fixing bacteria is to fix nitrogen, etc.

FUNCTIONAL ESSENTIALISM: drawing normative conclusions of any kind from descriptions of a thing's natural history.

5. MERITED IRRATIONALITY

I shall offer two arguments against FUNCTIONAL ESSENTIALISM: the first for the conclusion that being true is not the only biological function of belief (§6), and the second for the conclusion that belief has non-biological “functions” (§7). Both arguments are based on an empirical hypothesis about the ubiquity and advantages of certain forms of epistemic irrationality. This section sympathetically explains that hypothesis.

An early version of the hypothesis I have in mind was proposed by Hume in his *Treatise of Human Nature*. In a chapter on “greatness of mind,” he maintains that:

[N]othing is more useful to us in the conduct of life, than a due degree of pride, which makes us sensible of our own merit, and gives us a confidence and assurance in all our projects and enterprizes. [...] Fortune commonly favours the bold and enterprising; and nothing inspires us with more boldness than a good opinion of ourselves. Add to this, that tho' pride, or self-applause, be sometimes disagreeable to others, 'tis always agreeable to ourselves[.] Thus self-satisfaction and vanity may not only be allowable, but requisite in a character. (*Treatise*, III.iii.2)

Hume concludes that “a genuine and hearty pride or self-esteem ... is essential to the character of a man of honor.” (Ibid.) Now Hume's view seems to be that self-esteem is virtuous only when accurate, thus he calls “impertinent” that “almost universal propensity of men to over-value themselves,” (Ibid.) and says that “nothing can be more laudable, than to have a value for ourselves,” but only “where we really have the qualities that are valuable.” (Ibid.) Below we will have occasion to challenge this aspect of Hume's view. What is most notable in Hume's discussion, I think, is the fact that he defends self-esteem in explicitly non-epistemic terms, arguing that “[t]he merit of ... self-esteem is deriv'd from two circumstances, *viz.* its utility and its agreeableness to ourselves.” Self-esteem both “capacitates us for business” and “gives us an immediate satisfaction.” (Ibid.) Clearly these virtues might, at least in principle, be secured by epistemically irrational and false beliefs. Hume's insight in his discussion of self-esteem is to draw attention to the fact that living a good human life depends in part on how one thinks, and in particular on how one thinks about oneself: beliefs, or belief-forming faculties or practices, may have “merit” in virtue of their contribution to the quality of one's life, and this form of merit is conceptually distinct from epistemic virtue.

Hume's claim that self-esteem is essential to living well enjoys significant empirical support from research conducted by social psychologists in the last 40 years. And this research suggests that self-esteem is merited for more or less the reasons that Hume claimed it was. But many psychologists have maintained, *pave* Hume, that self-esteem is merited even when epistemically irrational. This is the hypothesis that I mentioned at the outset: certain forms of irrational believing are both ubiquitous (as Hume observed) but also advantageous for the believer. Following Shelly Taylor and Jonathan Brown's influential review article (1988, see

also their 1994 and Brown and Dutton 1995), we can catalog at least three distinct forms of merited irrationality:

Self-enhancement bias: Research suggests that people tend to overestimate their own positive qualities. Campbell 1986 found that subjects tended to rate themselves “above average” or “better than most,” by comparison with their peers, when it comes to a wide range of traits and abilities, including leadership ability, physical prowess, and attractiveness. Other studies (Lewishohn et al. 1980, Conway and Ross 1984, John and Robins 1994) compared subjects self-evaluations with those of disinterested observers, and found self-evaluations overestimated positive qualities. Thus the ubiquity of a “self-enhancement bias” – a tendency to overestimate one’s own positive qualities.¹²

But not only does self-enhancement appear ubiquitous, it appears to be in some important way advantageous to those who employ it, as moderately depressed subjects are one of the only groups of people who seem not to engage in cognitive self-enhancement. Lewishohn et al. 1980 found that depressed subjects’ self-evaluations did not diverge significantly from those of observers. Such research has led to the hypothesis of “depressive realism” (Abrahamson and Alloy 1981), that there is a correlation between depression and accurate views of the self. But it matters not whether this is true, or whether depressed subjects have irrationally negative views of themselves (as on the more traditional view, e.g. Beck 1967, and as urged by some critics of Taylor and Brown, e.g. Colvin and Block 1994). What matters, for my purposes here, is that this research suggests that normal, non-depressed people tend to have irrationally high opinions of themselves.

Irrational optimism: Research suggests that people tend to have unrealistic views of their own futures. Weinstein 1980 found that most subjects considered themselves less likely than others (in their peer group) to experience negative events (e.g. illness, accidents, divorce), and more likely to experience positive events. And, again, this is not true among subjects suffering from mild depression (Alloy and Ahrens 1987).

Illusion of control: Perhaps most famously, Langer 1975 and Langer and Roth 1975 found that subjects, in a competitive situation, readily believed that they have some degree of control over chance events. Langer 1975 found that subjects believed they had more control over the outcome of dice rolls when they threw the dice themselves, as opposed to when someone else was rolling. And, again, depressed subjects tended to have more realistic views of their degree of control (Alloy and Abrahamson 1979).

(For a review of all this – and references to dozens of additional studies on each of these phenomena – see Taylor and Brown 1988 and 1994, and Brown and Dutton 1995.)

So here is a rough statement of the hypothesis I’m proposing:

¹² For the sake of simplicity I here ignore several other phenomena that are standardly grouped under the heading of “self-enhancement bias,” including quicker recall of information that reflects well on the subject and the tendency to view positive traits, more than negative traits, as reflections of one’s true self.

MERITED IRRATIONAL SELF-ESTEEM: First, irrational self-esteem (in the form of a self-enhancement bias, irrational optimism, and an illusion of control) is ubiquitous, in the sense that it is statistically typical. In other words, most people are irrational in these ways. Second, irrational self-esteem is merited, in the sense that it advantageous, at least for most people, to employ self-enhancement, irrational optimism, and the illusion of control.

But what does “advantage” amount to here?

For my part, I think a case could be made that irrational self-esteem can be, and often is, part of the best human lives. In other words, I think a case could be made that the good human life can and often does involve irrational self-esteem. But for my purposes here I’ll propose a much more specific, and much weaker, claim:

BUSINESS: Irrational self-esteem (of the sort described) promotes a person’s motivation to complete ongoing projects (tasks, plans, actions), and thus promotes the subject’s ability to complete said projects (tasks, plans, actions).

As Hume would put it, irrational self-esteem “capacitates us for business.” There’s good evidence for this claim. Felson 1984 found that subjects with a positive self-image worked harder and longer on tasks. Bandura 1977 notes a correlation between people’s belief in their “personal efficacy” and their ability to complete tasks. More recently, a number of psychologists (Golwitzer and Kinney 1989, Taylor and Gollwitzer 1993, Armor and Taylor 2003, Kurman 2006) have concluded that irrational self-esteem is typical when (non-depressed) subjects are in action or working towards a goal, while such irrationality is diminished when such subjects are deciding what to do or how to complete a given task. More on this below.

The hypothesis I’ve described, and BUSINESS in particular, strikes many philosophers as counterintuitive. But it enjoys significant empirical support. Several things should be noted, however, by way of clarification. First, the idea here is that certain specific forms of irrationality are ubiquitous and advantageous. Self-enhancement involves “people’s beliefs about their stable, valenced attributes” (Brown and Dutton 1995, p. 1288), i.e. beliefs that concern one’s own personality, character traits, and abilities. People do not self-enhance by believing that they are fifteen feet tall. Taylor, et al. 1989 note that “negative information serves to keep illusions specific to appropriate areas of life and at adaptive levels.” (1989, p. 122) Irrational optimism (of the sort I’m talking about) involves people’s beliefs about their own futures, not about the future in general, or about other people’s futures (Regan et al. 1995). So the view I’ve described does not mean that one ought to be optimistic about global poverty, nor does it entail that “optimism,” in the everyday sense, is a virtue.

Second, many of the irrational beliefs in question do not admit of easy empirical disconfirmation; they concern “subjective” matters of character and personality (Taylor et al. 1989). In addition, most people are capable of incorporating evidence against their illusory beliefs without trauma or discomfort, both by being selective in the gathering and interpretation of evidence, and by subtle revision of their illusory beliefs. (See Taylor and Brown 1988, pp. 201-3, Taylor et al. 1989, and Janoff-Bulman 1989.)

Third, irrational self-esteem usually involves augmenting an existing quality, rather than believing in a non-existent one (Taylor and Brown 1994, p. 23). Unlike in Langer's original studies, the illusion of control is usually manifested in people believing that they have more control (than they really do) over events that they do in fact have some control over. To believe that you are responsible for the movement of the planets is a psychotic delusion; to believe that you are responsible for the academic job you landed is a comfortable bit of self-deception. (See also Baumeister 1989.)

The upshot of these three points is to show how merited irrationality remains "under control" – how such irrationality doesn't amount to being completely disconnected with reality, or with being constantly disappointed or shocked, or with simply being deluded. Given its ubiquity in non-depressed subjects, we have reason to think that such irrationality can be kept under control, in this sense. But these three points give us some indication of *how* merited irrationality can be kept under control, and thus remain merited.

6. FIRST ARGUMENT: BEING TRUE IS NOT THE ONLY BIOLOGICAL FUNCTION OF BELIEF

Assume that MERITED IRRATIONAL SELF-ESTEEM is true. I maintain that this speaks against the claim that being true is the unique biological function of belief. According to Millikan et al., a belief is performing its proper biological function only when it is true. But the beliefs described in §5 – beliefs resulting from a self-enhancement bias, instances of irrational optimism about one's future, instances of the irrational overestimation of one's degree of control over the world – are, I maintain, functioning properly, even if and when they are false (and irrationally formed, and not amounting to knowledge, etc.).

The fact that irrational self-esteem is ubiquitous is no proof that such beliefs are functioning properly. Failure to perform one's biological function can be ubiquitous, as noted in §3. What suggests the proper functioning of irrational self-esteem beliefs is, first, the advantages enjoyed by the believer, and, second, the fact that the cognitive mechanisms involved in irrational self-esteem have the appearance of having been designed to confer said benefits.

Consider the merits of irrational self-esteem when it comes to the believer's motivation to complete her projects (as in BUSINESS, above). On this basis, I think we can plausibly propose that one function of beliefs is to promote the motivation of the believer. Not only are irrational self-esteem beliefs advantageous, but the belief-forming mechanisms or practices involved have the appearance of having been designed to promote motivation. Gollwitzer and Kinney 1989 found that subjects were prone to "illusionary optimism" in beliefs about their degree of control when placed in an "implemental mind-set" – when they were working towards a goal already decided upon – but were not so prone when in a "deliberative mind-set" – when they were deciding what was to be done. This suggests that our habits of self-deception (when it comes to our beliefs about personal control) are fine-tuned: we overestimate our degree of control when it's useful to do so (when we need it for the end of motivation), but not when it would be imprudent (when we need to be realistic, to make informed decisions). Thus, doxastic self-promotion is selective and contextual, and appears to be sensitive to prudence. (See also Taylor and Gollwitzer 1993, Armor and Taylor 2003, and Kurman 2006.)

What would have to be the case for promoting motivation to be a biological function of belief? Given our Millikanian conception of biological functions (§4):

- (7) I have ancestors who had irrational self-esteem beliefs,
- (8) that said beliefs promoted the believer's motivation contributed to the fitness of said ancestors (relative to salient competitors), and
- (9) that I have beliefs is (at least partially) explained by (8).

Is this plausible? What I maintain is that it is no *less* plausible than (5) – (6), above. Given the evidence summarized in §5, we have just as much reason to think that our ancestors enjoyed evolutionary advantage in virtue of forming false beliefs which constituted their greater self-esteem, as we have to think that our ancestors enjoyed evolutionary advantage in virtue of forming true beliefs. The basis, after all, for our inclination to say that true beliefs are fitness-enhancing is just our speculative intuitions about things.

But more is needed, namely, a biological explanation of that fact that I have beliefs, based on the advantages of irrational self-esteem, as required by (9). Let's review our sketch (§3) of the explanation we would have to give in the case of (6). The idea there was that genes coding for truth-reliable cognitive mechanisms were historically selected for, in virtue of the fact that having many true beliefs contributed to the fitness of those ancestors of mine who enjoyed such cognitive mechanisms. What we would need to say, then, in defense of the claim that promoting motivation is a biological function of belief, is that genes coding for certain cognitive mechanisms, namely those responsible for irrational self-esteem, were historically selected for, in virtue of the fact that having increased motivation contributed to the fitness of those ancestors of mine who enjoyed such cognitive mechanisms. Is this plausible? And, more importantly, is it any less plausible than what is needed to vindicate the idea that being true is a function of belief?

Why do we think that having true beliefs contributed to the fitness of our ancestors? Perhaps we imagine something like this: two tribes of ancestors leave their caves, on the hunt for mammoth. One tribe, as a matter of genetics, has more reliable perceptual faculties than the other. The tribe with more reliable faculties quickly spots a mammoth, and accurately throw their spears into it. The mammoth runs into the forest, but the tribespeople give chase; their keen ears track the mammoth's bellows. Finally they catch it, bring the meat back to the cave, etc. Compare the tribe with less reliable faculties: they mistake a mossy rock for a mammoth, their spears miss their target, they bring nothing back to the cave, they starve and die, etc. The genes coding for more reliable faculties flourish; those coding for less reliable faculties do not.

However, it seems we can tell an equally plausible story on the side of promoting motivation. Suppose one of two tribes, with equivalent perceptual faculties, has, as a matter of genetics, a tendency for irrational self-esteem. If BUSINESS is right, we can imagine that this higher self-esteem tribe will put more effort into their hunt, be more willing to give chase to the mammoth, and so on. The tribe with lower self-esteem will be less likely to persevere. The result will be the same: the motivated hunters will bring home the food; the unmotivated will starve. Genes coding for irrational self-esteem will flourish; those not coding for irrational self-esteem will not.

How plausible is this? A crucial assumption is that irrational self-esteem is the result of genetically determined cognitive mechanisms. (Just as, above, we assumed that reliable perception is the result of genetically determined cognitive mechanisms.) Is this assumption plausible? I think the answer is: we don't know. But we also don't know, when it comes, in general, to the cognitive mechanisms responsible for reliable belief formation. As I said above (§4), defenders of functional essentialism have not offered the kind of genetic evidence that would be needed to make (6) anything more than a plausible hypothesis. (9) is no less plausible. In other words, we have just as much reason to say that promoting motivation is a biological function of belief, as we have for saying that being true is a biological function of belief. (And the same, *mutatis mutandis*, for any other advantages of irrational self-esteem.)

If all this is right, the correct analogue of belief is not the heart, but the liver. The liver has a plurality of functions, two of which are detoxification and glycogen storage. A properly functioning liver does both. So perhaps belief has (at least) two biological functions: (i) to representing reality accurately, by being true, and (ii) to promote the motivation of the believer, by representing her in a positive light. When a belief cannot do both (as often happens), a belief may fulfill one or the other of its proper function, perhaps depending on the costs or the importance of the proposition in question, or whatever. But just as detoxification does not have pride of place among the functions of the liver, neither does representing reality have pride of place among the functions of belief. If that is right, FUNCTIONAL ESSENTIALISM is false.

It could be objected that, although promoting motivation is one function of belief, it is not a function that flows from the essential nature of belief. Consider, again, the heart. To say that the function of the heart is to pump blood is consistent with a heart being used as a paperweight. A heart might even make a very good paperweight, but this will never make it the case that the function of the heart is to weigh down paper, since the function of a thing must flow from the essential nature of that thing. And it is not of the essential nature of a heart to be a paperweight. Or, perhaps less drastically, Velleman concedes that “[e]volution ... may have given us dispositions to ... overestimate our own popularity,” but that when an irrational belief is not responsive to evidence, “its regulative mechanisms are being prevented from doing what they were designed to do.” (Velleman 2000, p. 254, see also Wedgewood 2007, p. 165) But if evolution gave us certain dispositions to be irrational, in what sense are irrational beliefs, those that result from the very dispositions in question, being prevented from doing what they were designed to do? Perhaps such prevention is also “designed” – but in that case we simply have multiple functions of belief, which sometimes conflict.

7. SECOND ARGUMENT: BELIEF HAS NON-BIOLOGICAL “FUNCTIONS”

Assume that MERITED IRRATIONAL SELF-ESTEEM is true, and assume that we stick with the idea that being true is the unique biological function of belief. I then maintain that biological evaluation of beliefs is not the uniquely appropriate form of evaluation of beliefs. In other words, roughly, FUNCTIONAL ESSENTIALISM is no more plausible than the view that reproductive evaluation is the uniquely appropriate way to evaluate sex. We said above that there are a plurality of appropriate ways to evaluate sex, other than reproductive evaluation. This intuition results from the fact that, although the biological function of sex may very well be reproduction, there are many things that we care about other than performing our biological functions (or having our parts perform their biological functions). We could say

that there are non-biological “functions” of sex, given the fact that sex can do well vis-à-vis various things we care about other than reproduction, e.g. hedonistic pleasure, physical intimacy, being moral, and so on.

Similarly, we should say that there are non-biological “functions” of belief, given the fact that belief can do well vis-à-vis various things we care about other than epistemic goodness, e.g. subjective wellbeing and motivation to persist at projects. As we said above (§4), this could be formulated as a metaphysical point: the biological nature of belief does not determine the essential nature of belief, i.e. beliefs are not to be understood essentially as biological entities, and thus epistemic evaluation – assuming that this is identical to biological evaluation – is not the evaluation of beliefs, qua beliefs, and thus not the uniquely appropriate form of evaluation of beliefs. Or it could be formulated as a metaethical point: the evaluation of beliefs, qua beliefs, is not the uniquely appropriate form of evaluation of beliefs. It might have some special status if we had some unique kind of concern for having good beliefs, qua beliefs. But if we are convinced of MERITED IRRATIONAL SELF-ESTEEM, our zeal for good beliefs, qua beliefs, ought to diminish. And thus epistemic evaluation of beliefs will not be the uniquely appropriate form of evaluation of beliefs.

8. Conclusion

The arguments of §6 and §7 constitute a dilemma for FUNCTIONAL ESSENTIALISM, given the hypothesis of MERITED IRRATIONAL SELF-ESTEEM. Suppose, on the one hand, that irrational self-esteem is a biological phenomenon, the result of evolved mechanisms “designed” to promote the fitness of the believer. In that case, we cannot plausibly maintain that being true, or enjoying other positive epistemic statuses, is the unique biological function of belief. On the other hand, if irrational self-esteem is not a biological phenomenon, we should realize that we want more out of our beliefs than for them to do well from a biological point of view. Either way, we must reject the idea that, because of facts about the biological function of belief, epistemic norms are uniquely appropriate when it comes to evaluating beliefs.¹³

¹³ This material is part of a larger project; parts of that project (including parts of this paper) were presented in 2008 at the APA Central Division meeting, at the Arché Center at the University of St. Andrews, at the philosophy departments at Durham University, and at the University of Arizona, and in 2009 at a conference on Epistemic Goodness at the University of Oklahoma, and at a workshop on The Normativity of Belief and Epistemic Agency at the Instituto de Investigaciones Filosóficas, Universidad Nacional Autónoma de México. Thanks are due to my audiences at those talks as well as to Bryan Frances, Stephen Grimm, Brian Hazlett, Daniel Howard-Snyder, Uriah Kriegel, Robert Howell, Ted Sider, Nishi Shah, Ralph Wedgwood, and Dennis Whitcomb, and to my graduate assistant at Fordham, Xingming Hu.

BIBLIOGRAPHY:

- Abrahamson, L.Y., and Alloy, L.B. (1981), "Depression, non-depression, and cognitive illusions: A reply to Schwartz," *Journal of Experimental Psychology* 110, pp. 436-47.
- Alloy and Abramson (1979), "Judgment of Contingency in Depressed and Nondepressed Students: Sadder but Wiser?," *Journal of Experimental Psychology: General* 108, pp. 441-85.
- Alloy, L.B., and Ahrens, A.H. (1987), "Depression and pessimism for the future: Biased use of statistically relevant information in predictions for self versus others," *Journal of Personality and Social Psychology* 52, pp. 366-78.
- Armor, D.A., & Taylor, S.E. (2003), "The effects of mindset on behavior: Self-regulation in deliberative and implemental frames of mind," *Personality and Social Psychology Bulletin* 29, pp. 86-95.
- Bandura, A. (1977), *Social learning theory* (Prentice-Hall).
- Baumeister (1989), "The Optimal Margin of Illusion," *Journal of Social and Clinical Psychology* 8, pp. 176-89.
- Beck, A.T. (1967), *Depression: Clinical, experimental, and theoretical aspects* (Harper and Row).
- Boghossian, P. (2003), "The Normativity of Content," *Philosophical Issues* 13, pp. 31-45.
- (2005), "Is Meaning Normative?," in Nimtz and Beckermann (eds.), *Philosophy - Science - Scientific Philosophy* (Mentis), pp. 205-18.
- Brown and Dutton (1995), "Truth and Consequences: The Costs and Benefits of Accurate Self-Knowledge," *Personality and Social Psychology Bulletin* 21, 1288-1296.
- Campbell, J.D. (1986), "Similarity and uniqueness: The effects of attribute type, relevance, and individual differences in self-esteem and depression," *Journal of Personality and Social Psychology* 50, pp. 281-94.
- Colvin and Block (1994), "Do positive illusions foster mental health? An examination of the Taylor and Brown formulation," *Psychological Bulletin* 116, pp. 3-20.
- Conway and Ross (1984), "Getting what you want by revising what you had," *Journal of Personality and Social Psychology* 47, pp. 738-48.
- David, M. (2001), "Truth as the Epistemic Goal," in Steup (ed.), *Knowledge, Truth, and Duty: Essays on Epistemic Justification, Responsibility, and Virtue* (Oxford University Press), pp. 151-69.
- Davidson, D. (1983), "A Coherence Theory of Truth and Knowledge," reprinted in *Subjective, Intersubjective, Objective* (Oxford University Press, 2001), pp. 137-53.
- Dennett, D. (1971), "Intentional Systems," *Journal of Philosophy* 68, pp. 87-106.

- Felson, R.B. (1984), "The effect of self-appraisals of ability on academic performance," *Journal of Personality and Social Psychology* 47, pp. 944-52.
- Goldman, A. (1977), "Plain Sex," *Philosophy and Public Affairs* 6:3, pp. 267-287.
- Gollwitzer and Kinney (1989), "Effects of deliberative and implemental mind-sets on illusion of control," *Journal of Personality and Social Psychology* 56, pp. 531-42.
- Greco, J. (2003), "Knowledge and Credit for True Belief," in DePaul and Zagzebski (eds.), *Intellectual Virtue: Perspectives from Ethics and Epistemology* (Oxford University Press), pp. 111-34.
- Grimm, S. (2009), "Epistemic Goals and Epistemic Values," *Philosophy and Phenomenological Research* 77, pp. 725-44.
- Janoff-Bulman, R. (1989), "The benefits of illusion, the threat of disillusionment, and the limitations of inaccuracy," *Journal of Social and Clinical Psychology* 8, pp. 158-75.
- John and Robins (1994), "Accuracy and bias in self-perception: Individual differences in self-enhancement and the role of narcissism," *Journal of Personality and Social Psychology* 66, pp. 206-19.
- Kelly, T. (2003), "Epistemic Rationality as Instrumental Rationality: A Critique," *Philosophy and Phenomenological Research* 66, pp. 612-640.
- Kvanvig, J. (2003), *The Value of Knowledge and the Pursuit of Understanding* (Cambridge University Press).
- Kurman, J. (2006), "Self-Enhancement, Self-Regulation, and Self-Improvement Following Failures," *British Journal of Social Psychology*, 45, pp. 339-56.
- Langer, E.J. (1975), "The illusion of control," *Journal of Personality and Social Psychology* 32, pp. 311-28.
- Langer, E.J., and Roth, J. (1975), "Heads I win, tails it's chance: The illusion of control as a function of the sequence of outcomes in a purely chance task," *Journal of Personality and Social Psychology* 32, pp. 951-55.
- Lewishohn, Mischel, Chaplin, and Baron (1980), "Social competence and depression: The role of illusory self-perceptions," *Journal of Abnormal Psychology* 89, pp. 203-12.
- Millikan, R.G. (1984), *Language, Thought, and Other Biological Categories* (MIT Press).
- Nagel, "Sexual Perversion" (1979), in his *Mortal Questions* (Cambridge University Press), pp. 39-52.

Plantinga, A. "Introduction: The Evolutionary Argument Against Naturalism," in Beilby (ed.), *Naturalism Defeated? Essays on Plantinga's Evolutionary Argument Against Naturalism* (Cornell University Press, 2002), pp. 1-12.

Quine, W.V.O. (1969), "Natural Kinds," in *Ontological Relativity and Other Essays* (Columbia University Press), pp. 114-38.

Regan, Snyder, and Kassin (1995), "Unrealistic Optimism: Self-Enhancement or Person Positivity?," *Personality and Social Psychology Bulletin* 21, pp. 1073-82.

Riggs, W. (2002), "Reliability and the Value of Knowledge," *Philosophy and Phenomenological Research* 64, pp. 79-96.

----- (2009), "Two Problems of Easy Credit," *Synthese* 169, pp. 201-16.

Shah, N. (2003), "How Truth Governs Belief," *Philosophical Review* 112, pp. 447-82.

Shah, N., and Velleman, J.D. (2005), "Doxastic Deliberation," *Philosophical Review* 114, pp. 497-534.

Sosa, E. (2007), *A Virtue Epistemology: Apt Belief and Reflective Knowledge, Volume 1* (Oxford University Press).

----- (2009), "Knowing Full Well: The Normativity of Beliefs as Performances," *Philosophical Studies* 142, pp. 5-15.

----- (forthcoming), *Knowing Full Well: The Normativity of Beliefs as Performances*.

Taylor, S., and Brown, J., (1988), "Illusion and Well-Being: A Social Psychological Perspective on Mental Health," *Psychological Bulletin* 103, 193-210.

----- (1994), "Positive Illusions and Well-Being Revisited: Separating Fact from Fiction," *Psychological Bulletin* 116, 21-7.

Taylor, Collins, Skokan, and Aspinwall (1989), "Maintaining positive illusions in the face of negative information: Getting the facts without letting them get you," *Journal of Social and Clinical Psychology* 8, pp. 114-29.

Taylor and Gollwitzer (1993), "The effects of mindset on positive illusions," *Journal of Personality and Social Psychology* 69, pp. 213-36.

Velleman, J.D. (2000), *The Possibility of Practical Reason* (Oxford)

Wedgwood, R. (2002), "The Aim of Belief," *Philosophical Perspectives* 16, pp. 265-97

----- (2007), *The Nature of Normativity* (Oxford University Press).

Weinstein, N.D. (1980), "Unrealistic Optimism About Future Life Events," *Journal of Personality and Social Psychology* 39, pp. 806-20.