Naturalised Epistemic Oughts

Mona Simion

Cogito Epistemology Research Centre, University of Glasgow

**1. Introduction[[1]](#footnote-1)**

For a long time, epistemology has been negative epistemology: we cared about identifying restrictions on our permissions to believe, judge, use a proposition as a premise in reasoning, and so on. The assumption underlying this narrow focus has been that suspending is safe. Recent social phenomena - scientific evidence resistance, distrust in expertise, highly costly scepticism about well-established scientific claims such as vaccine safety and climate change - have painfully highlighted that this has been a mistake all along: suspending is just another doxastic attitude. There is nothing safe about it by default: it needs justification just as much as belief does. As a result, positive epistemology, discussing norms of evidence responsiveness and evidence gathering, as well as the effect that breaches thereof have on epistemic justification, is all the rage in the past several years.[[2]](#footnote-2)

Hilary Kornblith is a pioneer of positive epistemology: he has been talking about epistemic obligations to respond to and gather evidence, and their impact on epistemic justification, since at least as early as his 1983 ‘Justified Belief and Epistemically Responsible Action.’ Epistemology and policy alike would have been in a better place if the profession would have paid more attention to the far-reaching consequences for theory and practice of Kornblith’s results in this paper.

Kornblith’s work has been a major inspiration for my work on the source of epistemic normativity, evidence resistance, and epistemic oughts. Here are two claims that both Kornblith and I endorse:

1. Epistemic Oughts Affect Justification: Breaches of oughts of evidence responsiveness and evidence gathering affect one’s epistemic justification.
2. Epistemology Naturalised: Epistemic normativity is naturalistic normativity.

And here are two claims that Kornblith (e.g. 1983, 1993) endorses, but I don’t:

1. Epistemic justification implies epistemic responsibility.
2. Epistemic norms are desire-generated norms: they are hypothetical imperatives sourced in a desire for truth that we all ought to have, in virtue of its conduciveness to the satisfaction of any desire.

To be sure: Kornblith doesn’t take (3) and (4) to be strictly borne out by an epistemological picture that endorses (1) and (2) – i.e. by an account that is naturalistically respectable and incorporates epistemic oughts. However, it is fair to say that he takes (3) and (4) to fit snuggly into such a picture, and furthermore, to be strongly preferable to alternative accounts for one who likes such a picture.

 I argue here, *contra* Hornblith, that (3) and (4), far from snugly fitting in a framework that endorses (1) and (2), create problems for (1) and (2): the fan of naturalistic positive epistemology should stay away from desire-talk and responsibility-talk in explaining normativity. I then briefly sketch my preferred way of doing things with (1) and (2) and without (3) and (4).

**2. Epistemic Oughts**

Consider the following classic case of evidence resistance from Kornblith 1983:

**Dogmatic Physicist**: Jones is a headstrong young physicist, eager to hear the praise of his colleagues. After Jones reads a paper, a senior colleague presents an objection. Expecting praise and unable to tolerate criticism, Jones pays no attention to the objection; while the criticism is devastating, it fails to make any impact on Jones' beliefs because Jones has not even heard it (1983, 36).

Intuitively, Jones is not justified to continue believing the claims in his paper (at least not with the same degree of confidence): his colleague’s testimony constitutes a defeater for his justification. To put it in deontic terms, Jones (epistemically) ought to revise his doxastic attitude towards p (result in the paper) in the light of his colleague’s testimony, but fails to do so.

 Two questions arise for the naturalist epistemologist: (1) what is the (naturalistically respectable) normative source of this ‘ought’, and (2) by what mechanism does it affect Jones’s justification to believe? I will call these questions henceforth the Normative Source question and the Mechanism question.

*Two Elephants in the Room*

In order to make some quick progress towards answering these questions. let us start with the two elephants in the room:

First: The naturalist epistemologist is committed to thinking that philosophy should be continuous with the natural sciences, in several ways: the natural sciences should inform our philosophical theorizing, and our epistemological theories should be compatible with a worldview whereby nothing exists other than the sorts of objects, facts, properties, and relations studied by the natural sciences. The traditional normative challenge for the naturalist is to unpack epistemic normativity (and indeed, general normativity – but for our purposes here we will focus on the epistemic) in a naturalistically respectable fashion. That is, what is hard for the epistemological naturalist is to explain how such things as epistemic norms arise in nature. Note, crucially, that this challenge does not merely – as some have taken it to – amount to putting forth a reductive account of epistemic normativity in non-normative terms. More is needed: the view also needs to explain how come the purely descriptive property that normativity reduces to is something that *should* be instantiated. Reliabilism alone, for instance, does not meet this challenge: while reliabilism reduces epistemically normative phenomena (e.g. justification, defeat, knowledge) to purely descriptive phenomena – i.e. one variety or another of reliability – its champions still owe us an explanation as to why reliability is normative: why *should* our beliefs be formed via reliable processes, abilities etc.

Second: Recent work on the impact of violations of norms of evidence responsiveness and evidence gathering on epistemic justification has highlighted a new and harder challenge for the epistemic naturalist: the epistemic impermissibility of evidence resistance and bad inquiry suggests that there are such things as epistemic oughts governing our practices of belief forming, updating, and maintaining. As such, any epistemology that is able to predict epistemic impermissibility in cases involving failures to properly respond to and gather evidence will be an epistemology that is able to incorporate epistemic oughts. This is not trivial in general, and even less so for the naturalist. The first reason why it is hard is what the literature takes to be an in-principle difficulty, i.e. the very plausibility of the existence of these epistemic oughts: most people think voluntarism about belief is false – and indeed, naturalists have traditionally championed non-voluntarist positions. Even bold voluntarists would likely accept that, for the most part, we don’t exercise voluntary control over what we believe. Notably, though, many have thought that non-voluntarism is incompatible with epistemic oughts for belief: if ought implies can, the thought goes, and if I cannot believe at will, I cannot be subject to norms obliging me to do so either.[[3]](#footnote-3)

Note why this problem is especially hard for naturalists: naturalists are paradigmatically (and should be) non-ideal epistemologists: their accounts of justification, knowledge, etc. shy away from idealisation, and incorporate cognitive limitations. This way to proceed, however, constitutes a tacit acceptance of some (more or less restricted – see below) variety of ought-implies-can: we cannot be governed by norms that ask us to be logically omniscient and believe like a logically omniscient agent would, because we cannot be logically omniscient, nor can we believe like a logically omniscient agent would (Kornblith 2002). Similarly, the thought goes, since we cannot form beliefs at will, it seems worrisome to think that we ought to form them in some particular way.

A second and related version of this challenge has to do with the nature of these epistemic oughts: how should we conceive of the nature of epistemic oughts such that they are neither too normatively thick (such as to be in conflict with non-voluntarism), nor too normatively thin (such as to predict, against intuition, that epistemic normativity is shallow: that there is nothing substantively wrong with not respecting one’s evidence, believing based on wishful thinking, etc). I will call this henceforth *The Strength Dilemma* for the champion of epistemic oughts. Once more, this version of the challenge is harder for the naturalist than for the non-naturalist champion of epistemic oughts: that is because, whatever the answer to the question about the nature of these oughts on offer, it had better be naturalistically respectable as well – on top of avoiding the above dilemma.

Several pieces in the literature attempt to offer accounts of epistemic ought that bypass the voluntarism objection; I will not run through all of them in great detail here. To see the ought strength dilemma, let’s look at a couple of classic proposals.

According to Richard Feldman, epistemic obligations are what he calls ‘role oughts’:

There are oughts that result from one’s playing a certain role or having a certain position.

Teachers ought to explain things clearly. Parents ought to take care of their kids […]. Incompetent teachers, incapable parents […] may be unable to do what they ought to do. Similarly, I’d say, forming beliefs is something people do. That is, we form beliefs in response to our experiences of the world. Anyone engaged in this activity ought to do it right. In my view, what they ought to do is to follow their evidence (rather than their wishes or fears). I suggest that epistemic oughts are of this sort – they describe the right way to play a certain role. (Feldman 2000)

Feldman’s account is the paradigmatic case of a theory that is affected by the lack of ‘normative oomph’ worry: after all, role oughts can – and often are – in an important sense uninteresting, or even bad oughts. Role oughts generated by the role of mafia boss, for instance, seem too normatively thin to constitute the right kind of model for epistemic normativity.

Matthew Chrisman ventures to explain belief oughts in a non-voluntarist-friendly fashion by pointing out a classic distinction between norms governing actions, or ‘ought-to-dos’ – which may be subject to some variety of the ought-implies-can principle – and ‘rules of criticism’, governing states, or ‘ought-to-bes’, which, Chrisman thinks, are not thus constrained. Here is Chrisman:

In developing an account of robustly normative claims about what someone ought to believe, my strategy is to treat these as adverting to a species of state norms. For instance, the claim, ‘You ought to believe you are reading this text right now’ could be understood to be an instance of the general form, ‘S ought to be in doxastic attitude A towards proposition p under conditions C’. Then, the crucial observation is that some such normative claims seem to be true, just like ‘Clock chimes ought to be disposed to strike on the quarter hour’, ‘The beds ought to be made by 8am every morning’, ‘A child ought to be able to tie their shoes by age seven’ […]. Yet these true normative claims don’t presuppose that their subjects be capable of voluntarily following the rule. (Chrisman [2008](#ref_bib1_34))

The worry for Chrisman’s account is, once more, one having to do with ‘normative oomph’: state norms are clearly not always normatively thick – indeed, several of the examples given are paradigmatic examples of social standards of correctness. Is there any interestingly thick normative sense in which beds that are unmade by 8 a.m. are defective? Not really. Furthermore, most importantly, we want to say that the epistemic is in a completely different category of normative weight: I would worry about you if you formed unjustified beliefs on a regular basis in a completely different way than I would if I found out you’re not in the habit of making your bed in the morning. And the worry, intuitively, would be different in kind, not merely in strength.[[4]](#footnote-4) If so, we need to find extra normative resources to account for the epistemic.

**3. Desires for Truth and Epistemic Responsibility**

We need a model for epistemic normativity that, at the same time, circumvents voluntarism-related worries and is normatively thick enough to account for the intuitively non-conventional nature of the epistemic. Otherwise, if the model put forth is too normatively thin, the suspicion will be that it only circumvents voluntarist worries in virtue of its normative thinness – and thus that an extensionally more adequate, normatively thicker incarnation thereof will fail to do so.

Hilary Kornblith’s view of epistemic normativity makes great progress in this direction: according to Kornblith, in a nutshell, the fact that we all should have a desire for truth, together with the corresponding hypothetical imperatives it generates, is all we need for explaining epistemic oughts. We should all have a desire for truth because truth is conducive to all desire satisfaction: after all, if I desire food and shelter, I had better have some true beliefs about where and how to get them. That is all there is to epistemic normativity: not much more ‘normative oomph’ than this is needed to explain the data we need to explain. Here goes Kornblith:

Any account which does explain the source of epistemic normativity must explain how it is that epistemic claims have normative force. If you tell me that a belief of mine is unjustified, this gives me reason to give up that belief. The epistemic claim is something about which I should care, and an account of the source of epistemic norms must explain why it is that I should care about such things. Since having a desire for something gives one a reason to care about it, it is well worth considering whether desire might serve as the source of epistemic normativity (Kornblith 1993, 363).

According to Kornblith, epistemic norms are hypothetical imperatives towards the goal of believing truly. Crucially, true belief occupies a special place in the scheme of goals, which makes it such that epistemic norms – as dropping out of the goal of true belief – also occupy a special place in the scheme of hypothetical imperatives: truly believing is *universally* instrumentally valuable: no matter what goals you have, truly believing will best help you achieve those goals. Thus, for Kornblith, epistemic norms are hypothetical but universal: they apply to us as long as we possess any goal at all:

Precisely because our cognitive systems are required to perform evaluations relative to our many concerns, and to perform these evaluations accurately, the standards by which we evaluate these cognitive systems themselves must remain insulated from most of what we intrinsically value, whatever we may value. This provides a reason to care about the truth whatever we may otherwise care about. It also provides us with a reason to evaluate our cognitive systems by their conduciveness to truth. And this is precisely what epistemic evaluation is all about. Truth plays a pre-eminent role here (1993, 372).

This is Kornblith’s answer to the Normative Source question: desires are naturalistically decent entities – they are states of mind that aren’t in any way spooky in a fully naturalistic world view. In turn, the normative strength of epistemic norms is derived as follows: for all S, if S desires anything at all, then true beliefs have irreplaceable instrumental value for S. This makes it the case that conforming to epistemic norms is of irreplaceable instrumental value for S. This gives us all a reason to care about truth and epistemic norms (whether or not we realize this or act on it).

 How about the Mechanism question? How does this picture explain the mechanism by which epistemic oughts thus generated affect one’s justification? Kornblith’s answer (1983), in brief, is: responsible agents take the means to their ends. Since truth is a universal end (as per above), responsible agents obey epistemic norms. Here it goes:

An epistemically responsible agent desires to have true beliefs, and thus desires to have his beliefs produced by processes which lead to true beliefs; his actions are guided by these desires. […] when we ask whether an agent's belief is justified what we mean to ask is whether the belief is the product of epistemically responsible action, i.e. the product of action an epistemically responsible agent might have taken.[…] Being justified requires more than simply reasoning properly; it requires that one gather evidence properly as well (Kornblith 1983, 34-5).

According to Kornblith, then, our favourite physicist Jones is not justified to continue believing his results because his conduct is epistemically irresponsible; had Jones' actions been guided by a desire to have true beliefs, he would have listened carefully to the objection raised by his colleagues. Given that his maintaining his belief that his result holds is due, in part, to this epistemically irresponsible act, his belief is unjustified.

I share Kornblith’s naturalism, as well as, to a large extent, his moderate view on epistemic ‘normative oomph:’ epistemic normativity need not give us categoricity in the way moral normativity does. This way to think about things is perfectly compatible with the intuition that epistemic normativity is thick in the relevant sense – in that it is not a fleeting type of normativity, dependent on contextual whims.

I think, however, that a desire-based picture remains unsatisfactory on precisely the desideratum Kornblith and I both endorse: naturalistic friendliness. Three reasons for this:

*The Methodological Worry*

The first worry is methodological: recall that the methodological naturalist wants philosophical methodology to be continuous with scientific methodology. However, Kornblith’s proposal for the normative strength of the epistemic is largely armchair-substantiated: Kornblith takes it to be plausible from the armchair that most of our desires are best served by getting at the truth. While plausible from the armchair, this hypothesis seems to afford empirical investigation: imagine a possible world – indeed, some have argued that it may well be the actual world (Taylor & Brown 1988, Brown & Dutton 1995) – where getting things slightly wrong is best for us – i.e. for fulfilling our desires for general wellbeing, happiness, survival, etc. It is best for us to think we are slightly smarter and more beautiful than we already are, to harbour irrational optimism about what the future will bring, to think people around us have our best interest in mind, and to not bother too much with precise e.g. probability calculus or distance approximation, but rather to just work with numbers in the ballpark of truth. I don’t think this is the actual world: indeed, I think cases like these discussed in the literature are salient precisely because they are exceptional: for the most part, we are extremely reliable cognizers, and that is good for us. But the fact that Kornblith and I, from the armchair, agree on the plausibility of the thesis that this is how the actual world looks like tells us precious little about what a thorough epistemologically-informed empirical investigation would deliver.

*The Value-Theoretic Worry*

Enough about methodology: here is a second, more philosophically substantive reason to worry that Kornblith’s desire-based picture will fail him on precisely the naturalistic grounds that inspired it. Kornblith explains the epistemic normativity of belief in terms of its conduciveness to desire fulfilment. But now note that belief does not have exclusivity on normativity among mental states: desire is also, plausibly, normatively constrained (Schroeder 2020). There are such things as more and less fitting desires, and, indeed, desires that one should not have. If that is the case, by appealing to desire to explain the epistemic normativity of belief, Kornblith may just have pushed the normative burden from talk of one mental state to another: we are owed a naturalistic picture of the normativity of desire. Why think that the means for satisfying bad desires or desires you shouldn’t have, have any force on us, even if they are necessary for those desires? The answer that is readily delivered by Kornblith’s picture – that a truth desire is fitting because it underlies all other desires one might have –will not do, for two main reasons: first because, after all, these other desires may well, in many cases – or even overwhelmingly - be unfitting. So, once more, the normative explanation has been pushed further but not delivered: the fittingness of the desire for truth will now reduce to whatever makes these further desires fitting. And that, once more, requires a naturalistic normative explanation that we are not offered.

Furthermore, and relatedly, hypothetical imperatives are only as thick, normatively, as the desire triggering them is fitting, and a desire for truth will be just as fitting as the desires it is instrumental to fulfilling. Bad desires generate norms with little normative strength. Ideally, as naturalists, we want to explain this in a naturalistically friendly fashion: we want to be able to predict – in a naturalistic normative framework – why some norms are thicker than others – i.e., as it were, why nature cares about some desire-generated hypothetical imperatives more than about others -, and we want, ideally, to get the result that the epistemic is on the thicker side of things than norms generated by e.g. self-destructive desires, such as desires for smoking or drinking. In other words, whatever our framework, we should be able to predict that nature cares more about the epistemic than about smoking. Nature, plausibly, only cares about truth insofar as the desires it promotes are naturally beneficial. There is no naturalistic value to the true beliefs of the binge smoker and heavy drinker – since they are, to a large extent, supporting naturally unfitting desires.

Furthermore, even if we leave unfitting desires aside, Kornblith’s picture is facing a value swamping worry: the worry is that, on this instrumental picture, the value of truth will be swamped by the value of the other fulfilled desires. Recall Zagzebski’s (1996) coffee machine case: The value of reliable coffee producing machines is fully swamped by the value of the coffees produced: as soon as the coffee produced is good, there is no extra value to it also being produced by a reliable machine. Similarly, it would seem, the value of true belief is fully swamped by the value of the desires the fulfilment of which it supports: once the desires in question are fitting and fulfilled, it doesn’t matter if they were such in virtue of some true beliefs involved in the attempt at fulfilment or not.

*The Strength Worry*

Third, and finally, I worry about the naturalistic credentials of an account that explains the impact of epistemic oughts on epistemic justification in terms of epistemic responsibility. Two reasons for this: first, we know from ethics that there is such a thing as an epistemic condition on responsibility (Rudy-Hiller 2022): very roughly, where phi-ing is a token bad action, I am only responsible for phi-ing if either I was aware that I was doing something bad and did it anyway, or I should have been aware that I was doing something bad. But an account that purports to explain justification defeat in terms of failures of epistemic responsibility, since epistemic responsibility is further unpacked in terms of epistemic should, threatens either vicious circularity, or not being very informative, in that the ‘should’ in epistemic responsibility requires further unpacking.

Finally, an epistemic responsibility condition seems to be too strong a condition on epistemic justification: one way to unpack it that is straightforwardly too strong is by requiring the presence of a stable character trait. Kornblith alludes to something along these lines – but does not embrace it - when he writes the following about the case of physicist Jones:

As we might expect from the parallel with moral evaluation of actions, this attribution of an unjustified belief reflects ill on Jones' character, or at least that part of his character which plays a role in the retention of this belief. If this aspect of Jones' character continues to play a significant part in his actions, more and more of his beliefs will be unjustified, in spite of the fact that the reasoning which led him to these beliefs, and for which he retains them, remains impeccable (Kornblith 1983, 36).

However, clearly, a justified subject need not manifest great intellectual character – not even partly - , nor have it to begin with: the epistemically vicious among us can also form justified beliefs, even by employing otherwise vicious parts of their character. Donald Trump, climate change deniers, and conspiracy theorists have all kinds of justified beliefs: about the location of milk in the fridge, and the weather in Florida, but also about political matters: it’s not clear how we could isolate a ‘part’ of their intellectual character that does reliably well. Rather, what seems to be going on is that they get it right sometimes – maybe when motivated or otherwise sloppy reasoning doesn’t get in the way.

Conversely, and even more crucially for our particular purpose here, we need to be able to explain what is going wrong in cases of breaches of norms of evidence responsiveness and evidence gathering even if they feature exceptionally intellectually virtuous subjects: one-off cases of breaches of these norms are just as bad, epistemically, and they affect one’s justification just as much as cases in which the cognizer hosts an intellectually vicious character. We need, in other words, to be able to explain what goes wrong in one-off cases of evidence resistance, in which the subject’s bad epistemic character is not to blame – since the subject is a virtuous yet fallible believer.

My worry is that it will be difficult to spell out the responsibility condition even in weaker terms, in a manner that is permissive enough for allowing for ubiquity of epistemic justification but also efficient in dealing with resistance cases. To see why, note, once more, that we are fallible creatures: whatever virtuous character traits we may have, it will be consistent with them that we fail to manifest them on occasion, but also that we fail to do the right thing on occasion, in spite of manifesting them. No matter how courageous I might be, I may, on occasion, fail to be the first to reach the enemy’s lines. I need not do the perfectly courageous thing all of the time in order for my actions to be consistent with having (or manifesting) the relevant (fallible) character virtue. Furthermore, even in manifesting courage, I may end up doing things that look cowardly due to misreading the facts on the ground. Again, we are fallible creatures: in us, virtues can be manifested in instances of failure. If so, however, one-off cases of evidence resistance will be consistent with both having and manifesting great intellectual character traits.

One other option to spell out this responsibility condition would be to follow (Williamson forthcoming) and go modal: maybe what is required for the responsibility condition on epistemic justification to be met has to do with doing what the intellectually virtuous person would do in the situation at stake. Unfortunately, this will not work either, once more, in virtue of virtue fallibility: after all, the impeccably virtuous person involved in a one-off case of evidence resistance is doing what a virtuous person – namely, themselves – would do in that situation.

As far as I can tell, what is required to solve this problem is some variety of character trait infallibilism about evidence uptake: on a view like this, cases of resistance to evidence are not consistent with good epistemic character because the virtuous person just would not fail in this way. The view is not very appealing, I submit: as soon as we go down this route, we lose both general naturalistic plausibility for our account of epistemic character traits (we are fallible creatures!) and, even more crucially for our purposes, justification becomes a rare commodity, to be held only by the flawless among us.

**4. Functions and Norms**

I think that the view of epistemic normativity we are after should be naturalistic, exhibit prior normative plausibility, and be generalisable to other normative domains but also, at the same time, have enough ‘normative oomph’ to explain the intuitive categoricity of epistemic normative constraints. In this section, I will outline my view of the source of epistemic normative strength and argue that it meets all these desiderata: on this account, epistemic normativity is etiological functionalist normativity.

In traits, artefacts, and practices alike, functions generate norms (Millikan 1984). There is such a thing as a properly functioning heart, a properly functioning can opener, and a proper way to make coffee. If that is so, when we are interested in a particular type of norm governing a particular type of trait, it is helpful to first identify its function.

On the etiological theory of functions (Nender 1991, Godfrey-Smith 1994), functions turn on histories that explain why the item exists or operates in the way it does. Take my heart; plausibly, tokens of the type pumped blood in my ancestors. This was biologically beneficial for my ancestors’ survival, which explains why tokens of the type ‘heart’ continue to exist. As a result, my heart acquired the etiological biological function (henceforth also e-function) of pumping blood. Acquiring an etiological function (henceforth also e-function) is a success story: traits, artefacts, and actions get etiological functions of a particular type by producing the relevant type of benefit. My heart acquired a biological etiological function by generating biological benefit.

Importantly, while etiology does require some history of beneficial effects, it does not require an awful lot of it; what it all amounts to is explaining the existence/continuous existence of a trait through a longer or shorter history of positive feedback. Functions arise from consequence etiologies, etiologies that explain why something exists or continues to exist in terms of its consequences, because of a feedback mechanism that takes consequences as input and causes or sustains the item as output (Graham 2012, Simion 2016).

Functions can be of different sorts: there are biological functions, aesthetic functions, social functions, etc. In contrast to the orthodox Millikan view, my account takes functions to be typed by the corresponding benefit. As such, if a trait produces a benefit of type B in a system, the function thereby acquired will be a function of type B. The heart’s function to pump blood is a biological function in virtue of the fact that the produced benefit is also biological (i.e. survival). The function of art is an aesthetic function in virtue of the fact that the produced benefit is an aesthetic benefit. Now, of course, aesthetic benefit might, and often will, also result in biological benefit. This, however, in no way renders the function at stake a biological function. What is important to keep in mind is that the benefit that is essential to aesthetic function acquisition is the aesthetic one. The fact that biological benefit is also associated with the latter is a mere contingent matter of fact.

Here is, then, the full etiological account I like (2019, 2024):

**E-function**: A token of type T has the e-function of type B of producing effect E in system S iff (1) tokens of T produced E in the past, (2) producing E resulted in benefit of type B in S/S’s ancestors, and (3) producing E’s having B-benefitted S’s ancestors contributes to the explanation of why T exists in S.

Functions come with associated norms: these prescribe the right way to proceed in order to reliably enough fulfil the function in question under normal conditions. Because its function contributes to the explanation of its very existence, the trait in question ought to perform in a way that is associated with likely enough function fulfilment. Your heart will be properly functioning when it functions in the way that reliably enough delivers function fulfilment under normal conditions: it will beat at a rate between 40 and 100 beats/minute, which, under normal conditions (i.e. conditions similar to those present at the moment of function acquisition), reliably leads to pumping blood in your circulatory system.

In my view, generating knowledge is the main etiological epistemic function of our cognitive capacities. I have defended the knowledge function claim at length in the past (e.g. Simion 2019, 2024), and I will not rehearse all these arguments here. Here are, in brief, a couple of reasons in support of the claim: First, knowledge is *an* etiological function of our cognitive capacities in that it meets E-Function: knowledge has been generated by our cognitive capacities and those of our ancestors in the past (after all, knowledge is readily available in our environment), that this benefitted our organisms (e.g. by informing us about the presence of predators and the whereabouts of food), and that the fact that knowledge benefitted us in this way contributes to the explanation of why cognitive capacities continue to exist in individuals like us.

Second, I have argued that generating knowledge is the *main* *epistemic* etiological function of our cognitive capacities. Two reasons for this: first, for all types t, main t-function attributions are value laden: the most valuable t-function is the main t-function. The function of your heart is pumping blood in your circulatory system. It is not just pumping, nor is it pumping blood in your circulatory system and making a ticking sound. Value matters for main functions. But it is widely accepted[[5]](#footnote-5) that knowledge is the most valuable epistemic state. If so, knowledge is the main etiological epistemic function of our cognitive capacities. Second, main functions tend to map on to attributive goodness (Geach 1956): e.g. the main practical function of the knife is to cut – and what it is for a knife to be a good knife simpliciter – i.e. a good token of its type – is for it to be a sharp knife. Many have argued that good belief – belief that is good as a token of its type – is knowledgeable belief (Williamson 2000). If so, the main etiological epistemic function of our belief formation capacities is to generate knowledge.

More things can be said on this topic than space permits. What is of central interest for us here is that the knowledge function generates norms: our cognitive capacities are properly functioning just in case they work in a way that is normally conducive to generating knowledge. When that happens, the beliefs they generate are justified. I dub my view of justification ‘knowledge-first functionalism’: the account is functionalist in that it takes the epistemic normativity of belief to drop out of the etiological epistemic function of our cognitive capacities. It is knowledge-first epistemological in that, unlike traditional, truth-first functionalism, it unpacks the function at stake in terms of knowledge. Here is the view:

**Knowledge-first functionalism (KFF):** A belief is prima facie justified if and only if it is generated by a properly functioning cognitive capacity that has the etiological function of generating knowledge (Simion 2019, 2024).

On this knowledge-centric picture, good belief is knowledgeable belief, while justified belief – belief that is permissible by the epistemic norm of belief – is belief generated by a properly functioning cognitive capacity that has the etiological function of generating knowledge. The standards for proper functioning are thus natural normative standards, and they are constitutively associated with promoting knowledgeable beliefs. Compatibly, one can be justified in holding false beliefs: proper function need not imply function fulfilment. Our cognitive capacities with the function of generating knowledge may and likely will fail to do so in abnormal conditions.

What is the source of epistemic obligation? What grounds the epistemic ought, in my view, is, once more, proper epistemic functioning. Our belief-formation capacities can malfunction in at least two input-dependent ways: via taking up the wrong kind of inputs (e.g. wishes), but also, and crucially for my purposes here, via failing to take up easily available good inputs (i.e. easily available evidence). It is overwhelmingly plausible that proper evidence uptake is constitutive of proper cognitive functioning: after all, plausibly, proper evidence uptake is involved in the manner in which our cognitive capacities generate knowledge in normal conditions. Pieces of evidence are pro tanto, prima facie justification-makers: they are the proper inputs to our processes of belief formation. When our belief-formation processes either fail to take up justifiers that they could have easily taken up or they take them up but fail to update accordingly, they are malfunctioning. The proper function of belief-formation capacities is input dependent: failing to take up the right inputs – whether this occurs by taking up the wrong inputs or by failing to take up the right inputs – is an instance of malfunctioning.

It is important to note that empirical results also overwhelmingly confirm the hypothesis that the proper functioning of our cognitive systems is input dependent (i.e. that our cognitive systems are malfunctioning if they fail to respond to environmental stimuli). Our levels of neuroplasticity – the brain’s disposition for neuron-level changes in response to the environment – predict the brain’s capacity to take up information from the environment: when a cognitive system displays abnormally low levels of structural neuroplasticity, learning in response to novel stimuli from the environment fails to occur at a normal rate. In turn, abnormally low levels of neuroplasticity, generating low responses to environmental stimuli, predict improper cognitive functioning. But if this is so, the proper functioning of our cognitive capacities is input dependent: one way in which they can malfunction is by failing to respond to easily available environmental stimuli.

Of course, there are limits to our trigger-responsiveness (Simion 2024), since there are quantitative limitations on our information accessing and processing: the fact that there’s a table somewhere towards the periphery of our visual field – in contrast with it being right in front of us, in plain view – is not something we can easily take up and process: we lack the power to process everything in our visual field, it’s just too much information. Similarly, while we might easily access any of *p*1, …, *pn* independently, we might not be able to easily access them at the same time, nor their conjunction, nor what follows from them, due to our processing limitations. This explains why we are not required to form a large number of beliefs at the same time, believe very long conjunctions, or form beliefs about all arithmetical truths, and why facts that we are not able to know cannot constitute evidence we should be responsive to.[[6]](#footnote-6)

 Quantitative limitations will deliver an epistemic obligation to take up only a proper subset of available facts *p*1, …, *pn* that is as large as your quantitative take-up limitations. Availability rankings will deliver the relevant set: the most easily available subset of facts that you can take up is the subset you should take up.

Besides quantitative limitations, our ability to respond to triggers that is at stake here will be further restricted by *qualitative* features: there are of *types* information that we just cannot access or process, and *types* of support relations that we cannot access or process. Finally, there are also *environmental limitations* – i.e. features of the social and physical environment that limit availability. In sum, the relevant limitations will track a ‘can’ for a typical cognizer of the sort exemplified.

We now have a straightforward explanation of what goes wrong in cases like Kornblith’s DOGMATIC PHYSICIST: it is an epistemic incarnation of input-level malfunction.

What about cases in which evidence is not readily available to the cogniser, but it *should* have been? Consider:

THE LAZY DOCTOR. Doctor Smith, unlike physicist Jones, is close to retirement and not very excited about medicine anymore. He believes that (p =) stomach ulcer is caused by stress, a proposition that was once widely believed in his field. As a result of laziness, X has failed to keep an up-to-date understanding of the literature. One of the many things they have missed is a well-known development (q =) that ulcer is caused by a bacteria (adaped from Goldberg, 2018).

In the case of Doctor Smith, one might think, as opposed to the case physicist Jones, the failure at stake is not one of evidence responsiveness, but rather one of evidence gathering. Aren’t these cases going to be better explained by a framework such as Kornblith’s (1983), on which epistemic responsibility is part and parcel of epistemic justification[[7]](#footnote-7), than by the functionalist framework? After all, lazy doctor Smith seems to be epistemically irresponsible rather than not-properly-trigger-responsive.

 I want to resist this intuition: available evidence will do all the work that we need done here. To see this, compare the following two variations on LAZY DOCTOR:

THE LAZY DOCTOR\*. Doctor Smith\* believes that (p =) stomach ulcer is caused by stress, a proposition that was once widely believed in his field (but which, unbeknownst to him, recent studies suggest to be false). The evidence that he bases his belief on - i.e. the literature he read a while ago - strongly supports this. Smith\* has no evidence that there are new studies challenging this view: everyone he has talked to about this issue – his colleague George, his boss Mary – has confirmed that the stress hypothesis stands. Furthermore, Smith\* is a doctor in a remote village, with no internet access, where all the colleagues he trusts – and that proved reliable and sincere in the past on all kinds of topics – told him there are no updates on the cause of ulcer.

THE LAZY DOCTOR\*\* Doctor Smith\*\* believes that (p =) stomach ulcer is caused by stress, a proposition that was once widely believed in his field (but which, unbeknownst to him, recent studies suggest to be false). The evidence that he bases his belief on - i.e. the literature he read a while ago - strongly supports this. Smith\*\* is a doctor in a big city, where he has access to plenty of available evidence that new studies have been carried on the topic. Nevertheless, Smith doesn’t care enough about the issue to investigate further.

I submit that Smith\*, but not Smith\*\*, is justified to believe that stomach ulcer is caused by stress: he relies on reliable (albeit not infallible) testimony that provides him with evidence that ulcer is caused by stress. In contrast, Smith\*\* should have inquired better, given the evidence that he has. But note that the salient difference between Smith\* and Smith\*\* is one in evidential situation. Smith\*\* could have done better, given his evidence: as opposed to Smith\*, he is in a position to know that he should read up on new developments in the field. The difference in justification, once more, boils down to a difference in available evidence.

Let us now go back to the strength dilemma generated by voluntarist worry and the question of ‘normative oomph’: the good news is that, clearly, norms of proper function do not imply any voluntarist claim: just like my heart is governed by norms of proper function about blood pumping, my cognitive capacities are governed by norms of proper function about belief forming.

Is functionalist normativity going to generate a normatively thick enough model for epistemic normativity? Let’s go back to norms and practices in the domain of mafia. Here is a worry one might have: mafia practices may well continue to exist because they achieve the corresponding ‘values’ internal to the domain of mafia. If so, on a functionalist picture, we can get norms out of these functions: norms that regulate proper ways of going about achieving the domain-specific values that the domain of mafia is organised around. But do the resultant ‘norms’ have any normative oomph at all?

One would think that the case of epistemic norms is different. For instance, the fact that S said that p and that S is very reliable on p-related matters seems like a reason to believe that p has normative force – I can’t just set it aside in the way I can set aside the orders from my crime boss. These cases case feel different in a way that needs to be explained.

On functionalism, indeed, it can be the case that, for example, efficient killing is a domain-specific value in the domain of mafia, which, in turn, generates corresponding (domain-internal) functionalist norms. It’s a completely different question, however, if the domain itself is, for example, valuable simpliciter to begin with – and I take it that the domain of mafia is not. If so, the normative force of its norms will be restricted to the domain of mafia.

Note also that evolved functions will also not safeguard against this worry: the originally selected biological function of a capacity need not always continue to serve the organism well, through changing environments; vestigial selected capacities can be outright harmful to the organism.

I take it to be empirically plausible that doing well epistemically is, at least for the most part, good for us, biologically. If this hypothesis is true, the domain of the epistemic, in contrast to the domain of mafia, will itself be valuable, and its internal functional norms will bear ‘normative oomph’ in a way in which norms of the domain of mafia do not. Importantly, though, I can afford to stay neutral on the extent to which the epistemic is good or bad: for ‘normative oomph’, I just need the fairly weak claim that it is generally good for us. This is important because it gives my view the flexibility to adapt to empirical results that purport to show that, at times, epistemic unreliability co-varies with biological benefits.

Humans are not fully bound by selected goals (Stanovich 2005): we often pursue our own (good or bad) goals, that don’t do much to benefit species survival. It may well be, thus, that we will continue knowledge searching even in times when it will stop being beneficial for us. It is compatible with my account that there comes a time when having lots of knowledge is harmful to us: under such circumstances, I am happy to give up the normative oomph claim – and I think that this is a feature of the view, not a bug: likely, the oomphiness intuition will follow suit and disappear as well.

Compatibly with all this, plausibly, the domain of epistemology is (for now) valuable for our survival in a way that generates thick normative constraints. My functionalism thus does have the capacity to distinguish different kinds and strengths of normative force. On my view, epistemic norms have (1) domain-bound normative force, in that they promote knowledge, which is the value around which the domain is organised, and (2) non-domain-bound normative force, in that ‘the epistemic’ is a domain that is (empirically plausibly) valuable for our survival.

**5. Conclusion**

Hilary Kornblith’s pioneering work in naturalizing epistemic oughts has triggered major advances in work on positive epistemology. I have argued here that, building on his work, we can make better progress as naturalists if we abandon desire-generated norms and epistemic responsibility-infused justification: far from snugly fitting in a naturalistic framework that both Kornblith and I endorse, these theses create trouble for the naturalist. Or so I have claimed.

We are better off doing things with functions and functionalist norms. The main etiological epistemic function of our cognitive capacities is to generate knowledge, and justification turns on this function: we are justified to believe just in case our cognitive systems work in the way in which they generate knowledge reliably under normal conditions. In turn, functioning in this way implies proper evidential uptake.

Finally: note that taking up the functionalist commitments defended here should be easy for naturalists of any sort – and indeed, for any epistemologist. Here is why: The fact that our cognitive system has an epistemic function is something that anyone that is minimally conceptually competent with the concepts involved should accept (even people who, e.g., might think that the epistemic function is not the only function, or not even the main function of our cognitive capacities). But if something has a function of type T to phi, then it can T-malfunction or T-function properly with respect to phi-ing. Malfunctioning and proper functioning are normative: traits T-ought not function in the way in which they do when they T-malfunction; rather, they T-ought to function in the way in which they do when they T-properly function. In this way, functionalist epistemic oughts should be accepted by anyone who accepts that our cognitive system has an epistemic function.

**References**

Brown, J. and Dutton, K. (1995). Truth and consequences: the costs and benefits of accurate self-knowldege. *Journal of Personality and Social Psychology*, 21: 1288–1296.

Chrisman, M. (2008). Ought to Believe. *Journal of Philosophy*, 105/7: 346-370.

Feldman, R. (2000). Voluntary Belief and Epistemic Evaluation. In Knowledge, Truth, and Duty, ed. M. Steup, Oxford: Oxford University Press: 83-97.

Flores, C. and Woodard, E. (2023), ‘Epistemic norms on evidence- gathering’, *Philosophical Studies*, 180 (9): 2547-2571.

Friedman, J. (2017). Inquiry and Belief. *Noûs* 53 (2):296-315.

Geach, P. (1956). Good and evil. *Analysis*, *17*, 32–42.

Godfrey-Smith, P. (1994). A Modern History Theory of Functions. *Noûs* 28: 344–62.

Goldberg, S. (2016). On the Epistemic Significance of Evidence You Should Have Had. Episteme 13 (4):449-470.

Goldberg, S. (2018). To the Best of Our Knowledge: Social Expectations and Epistemic Normativity. Oxford: Oxford University Press.

Graham, P. (2012). Epistemic Entitlement. *Nous* 46/3: 449-482.

Ichikawa, J. (2024). *Epistemic Courag*e. Oxford: Oxford University Press.

Kelp, C. (2021). *Inquiry, Knowledge, and Understanding*. Oxford: Oxford University Press.

Kelp, C. and Simion, M. (2017). Commodious Knowledge. *Synthese,* 194: 1487–1502.

Kornblith, H. (1983). Justified Belief and Epistemically Responsible Action. The Philosophical Review, Vol. 92/1, pp. 33-48.

Kornblith, H. (1993). Epistemic Normativity. *Synthese*, Vol. 94/3, pp. 357-376.

Kornblith, H. (2017). Doxastic Justification is Fundamental. *Philosophical Topics*, Vol.45, No. 1, pp. 63-80.

Kornblith, H. (2022). What does Logic have to do with Justified Belief? In Silva, P. and Oliveira, L. (eds.), *Propositional and Doxastic Justification: New Essays on their Nature and Significance*. New York: Routledge.

Millikan, R. (1984). *Language, Thought, and Other Biological Categories*. Cambridge/MA: MIT Press.

Neander, K. (1991). Functions as Selected Effects: the Conceptual Analyst’s Defence. *Philosophy of Science* 58: 168–84.

Oliveira, L. (2020). Evading the Doxastic Puzzle by Deflating Epistemic Normativity. In Stappleford, S. and McCain, K. (eds.). Epistemic Duties: New Arguments, New Angles. New York: Routledge.

Pritchard, D., Turri, J. and Carter, J.A. (2022). "The Value of Knowledge", *The Stanford Encyclopedia of Philosophy*(Fall 2022 Edition), Edward N. Zalta & Uri Nodelman (eds.), URL = <https://plato.stanford.edu/archives/fall2022/entries/knowledge-value/>.

Rudy-Hiller, F. (2022). "The Epistemic Condition for Moral Responsibility", *The Stanford Encyclopedia of Philosophy*(Winter 2022 Edition), Edward N. Zalta & Uri Nodelman (eds.), URL = <https://plato.stanford.edu/archives/win2022/entries/moral-responsibility-epistemic/>.

Ryan, S. (2003). Doxastic Compatibilism and the Ethics of Belief. *Philosophical Studies* Vol. 114/1/2: 47-79.

Schroeder, T. (2020). "Desire", *The Stanford Encyclopedia of Philosophy*(Summer 2020 Edition), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/sum2020/entries/desire/>.

Simion, M. (2024). *Resistance to Evidence*. Cambridge: Cambridge University Press.

Simion, M. (2019). Knowledge-First Functionalism. *Philosophical Issues*, 29: 254–267.

Simion, M. (2016). Perception, History and Benefit. *Episteme,* 13: 61–76.

Simion, M. and Kelp, C. (2018b). How to Be an Anti-Reductionist. *Synthese.* Online First.

Taylor, S. E., and Brown, J. D. (1988). Illusion and well-being: a social psychological perspective on mental health. Psychological Bulletin, 103(2), 193.

Willard-Kyle, C. (forthcoming). The Knowledge Norm for Inquiry. *Journal of Philosophy*.

Williamson, T. (2000). Knowledge and Its Limits. Oxford University

1. Many thanks to Jennifer Nagel and Luis Oliveira for extremely useful feedback on an early version of this paper. This research has received funding from the European Research Council (ERC) under the European Union’s Horizon 2020 research and innovation programme (KnowledgeLab: Knowledge-First Social Epistemology project, grant agreement No 948356).  [↑](#footnote-ref-1)
2. E.g. Friedman 2017, Kelp 2021, Goldberg 2018, Ichikawa 2024, Simion 2024, Willard-Kyle 2023, Woodard and Flores 2023. [↑](#footnote-ref-2)
3. See e.g. Ryan 2003 for an overview as well as excellent arguments against the voluntarism puzzle. [↑](#footnote-ref-3)
4. But see e.g. Oliveira 2020 for a proposal that embraces thin epistemic oughts. [↑](#footnote-ref-4)
5. See e.g. Kelp and Simion 2017 for a defence, and Pritchard, Turri, and Carter 2022 for an excellent overview of the debate. [↑](#footnote-ref-5)
6. See (Williamson 2000) for more on this. [↑](#footnote-ref-6)
7. In more recent work, Kornblith (2017, 2022) argues that, given human psychological limitations, doxastic justification, rather than propositional justification, is the fundamental notion in epistemology, and that we should understand the latter in terms of the former. I am sceptical on two grounds: first, as I have argued in the past, views that follow this recommendation traditionally struggle with cases of evidence resistance (Simion 2024). Second, the step from human psychological limitations restricting epistemic justification to putting doxastic justification first is too quick: a view that unpacks propositional justification in accordance to human psychological limitations, and doxastic justification in terms of it, remains available (and preferable – for one, for its performance on cases of evidence resistance). See (Simion 2024) for such a view. [↑](#footnote-ref-7)