

Normal Knowledge

Toward an explanation based theory of knowledge

Andrew Peet & Eli Pitcovski

Abstract

In this paper we argue that knowledge is *characteristically safe* true belief. We argue that an adequate approach to epistemic luck must not be indexed to methods of belief formation, but rather to explanations for belief. This shift is problematic for several prominent approaches to the theory of knowledge, including virtue reliabilism and proper functionalism (as normally conceived). The view that knowledge is characteristically safe true belief is better able to accommodate the shift in question.

Introduction

In this paper, we present a problem for theories of knowledge which place belief forming processes at center stage. We do so by arguing that adequate approaches to epistemic luck, and what we call the “non-deviance” condition on knowledge, must be given in terms of explanations for belief rather than belief forming methods or processes.

More precisely, our argument runs as follows: We start with the assumption that a belief constitutes knowledge iff it satisfies the following three conditions:

TRUTH: The belief is true.

ANTI-LUCK: The belief is not merely luckily true.

NON-DEVIANCE: The belief is held, and true, in a non-deviant way.¹

We briefly illustrate ANTI-LUCK and NON-DEVIANCE. We then provide a case in which the three conditions are satisfied, despite the agent’s belief being unsafe at the time of formation. We conclude that it is a case of unreliably formed knowledge.² In order to deal with such cases we need an anti-luck principle which is indexed to explanations rather than methods. It is shown that, given their take on ANTI-LUCK and NON-DEVIANCE, several prominent approaches to the theory of knowledge (such as proper functionalism and virtue reliabilism) are undermined by the shift to explanation indexing. We close by suggesting an alternative: that knowledge is *characteristically safe* true belief. We will begin by briefly illustrating ANTI-LUCK and NON-DEVIANCE.

The Anti-Luck Assumption

The need to rule out luck is illustrated by Gettier cases. Consider the following case (paraphrased from Gettier (1963)):

¹ We intend the non-deviance assumption to imply that, if it is to constitute knowledge, a belief must be justified. However, as we will explain, a belief can be justified and still be true in a deviant way.

² Fricker (2016) and Turri (2013) also argue for unreliably formed knowledge. Our argument is distinctive in that we do not advocate unreliable knowledge, merely *unreliably formed* knowledge.

GETTIER: Smith and Jones are competing for the same job. Smith has good reason to believe the following:

[a]. Jones will get the job, and Jones has ten coins in his pocket.

He infers the following:

[b]. The man who will get the job has ten coins in his pocket.

Smith's belief is justified and, as it happens, true. But it is only true because, unbeknownst to Smith, he, and not Jones, will get the job. Moreover, he has ten coins in his pocket.

Despite having a justified true belief Smith fails to acquire knowledge. Thus, knowledge is not justified true belief. Central to GETTIER is the fact that Smith's belief is accidentally true; it is only true as a matter of luck.

To handle Gettier cases, a satisfactory approach to knowledge must have the resources to rule out epistemic luck. The most prominent anti-luck condition, and the one to which we are most sympathetic, is the safety principle.³ According to a rough formulation of safety, a belief is safe iff it could not easily have been false. So, if an agent's belief that p is to constitute knowledge then there cannot be any nearby possibilities in which the agent has a false belief that p . This crude version of the safety principle is inadequate. It gets the wrong result in a number of cases. Consider, for example, the following case (based on a case presented by Nozick (1981)):

OLD LADY: Martha is an old lady. She has an adult son who is frequently ill, often on the verge of death. Her son comes to visit her and, upon seeing him standing in front of her, she forms the belief that her son is alive. However, if her son was dead then her daughter in law would have told her that he was alive. So Martha would have believed he was alive even if he was dead.

Given that Martha has seen her son standing right in front of her it is not a matter of luck that she has a true belief. She knows he is alive. Nonetheless, there are nearby possibilities in which her son is dead and she still believes he is alive.

This problem can be dealt with by indexing safety to some additional factor: by holding some factor fixed across nearby worlds and only considering worlds where that factor remains invariant. The standard strategy is to index safety to the method of belief formation. If we only consider worlds in which Martha forms her belief by seeing her son, then we also rule out the nearby worlds in which she has a false belief about whether her son is alive. This more sophisticated safety principle is presented by Pritchard (2005) as follows:

If a believer knows that p , then in nearly all, if not all, nearby possible worlds in which the believer forms the belief that p in the same way as she does in the actual world, that belief is true. Pritchard (2005), 163.

³ The safety principle is not universally accepted. It has been objected to by Neta and Rohrbaugh (2004), Come-saña (2005), Lackey (2006, 2008), Bogardus (2014), and Goldberg (2015). Much of what we say will apply to alternative modal anti-luck conditions such as the sensitivity condition. This is because any such approach will involve some form of indexing (such as indexing to methods).

We agree that safety must be indexed. However, we are not convinced that safety must be indexed to methods of belief formation. We will later argue that there are cases in which an agent can come to know despite having formed their belief via an unsafe method. By indexing safety to the explanation for the agent's holding their belief at a time we can make the correct predictions about such cases. Thus, we believe that safety should be indexed to explanations.⁴ Before presenting this argument let us quickly outline NON-DEVIANCE.

The Non-Deviance Assumption

According to NON-DEVIANCE, if an agent possesses knowledge then their non-luckily true belief must be explicable in a non-deviant way. This assumption is inspired by what Greco (1999) calls 'the problem of strange and fleeting processes': the problem that certain belief forming processes, despite being reliable (producing safe and true beliefs), do not produce knowledge, as they are somehow deviant.

The operative notion of 'deviance', unlike the notion of luck, does not exist as clearly within our pretheoretic conceptual repertoire. But the idea should be clear enough to those familiar with reliabilist approaches to knowledge, especially virtue reliabilism and proper functionalism. Such theories place restrictions on the sorts of processes which can give rise to knowledge. The proper functionalist holds that knowledge must be the output of a properly functioning process.⁵ That is, a (usually cognitive) process which functions, in normal conditions, to produce true beliefs. Any belief which is not the output of such a process will be held in a deviant way. The virtue reliabilist holds that knowledge is a cognitive achievement: if an agent has knowledge then their true belief is the result of the exercise of their cognitive abilities. If the agent does not believe truly because of their cognitive abilities then their belief is deviantly formed or deviantly true, and thus fails to constitute knowledge.⁶ It will be helpful to consider some examples in which deviance precludes knowledge.

One way for a belief to be deviant is for it to be deviantly formed; for it to be unjustified. Greco (1999) illustrates the problem with the following example (from Plantinga (1993)):

BRAIN: Sandy has a brain lesion. He has not visited a doctor, nor has he experienced headaches or any of the other symptoms we would normally associate with brain lesions. Nonetheless, he believes that he has a brain lesion. This is because he has an unusual type of brain lesion which causes its victims to believe they have a brain lesion, even if they lack any additional evidence that they have a brain lesion.

Sandy's belief is true, and it is safe. Nonetheless, he does not know about his brain lesion. The process by which he formed his belief is not of the right kind to produce knowledge. Virtue reliabilists and proper functionalists capture this by pointing out that Sandy's true belief is neither the result of his epistemic abilities, nor the output of a properly functioning belief forming process (respectively).

⁴ This will usually give the same result as method indexing because, in most cases, the method by which an agent acquires their belief, together with the relevant background conditions, will explain why they hold the belief. For example, the explanation for the old lady's belief is the fact that she saw her son. So, given the explanation for her belief, she could not easily have falsely believed that he was alive.

⁵ See Plantinga (1993) and Graham (2016, forthcoming).

⁶ Examples of virtue reliabilism include Greco (1993, 2003, 2007, 2010), Sosa (2007, 2010), Pritchard (2012), Kelp (2013), and Broncano-Berrocal (forthcoming).

NON-DEVIANCE does not simply reduce to the assumption that justification is necessary for knowledge. The explanation for a belief's truth or safety can be deviant despite the belief itself being justified. Consider the following case from Pritchard (2012):

TEMP: Temp forms his beliefs about the temperature in the room by consulting a thermometer. His beliefs, so formed, are highly reliable, in that any belief he forms on this basis will always be correct. Moreover, he has no reason for thinking that there is anything amiss with his thermometer. But the thermometer is in fact broken, and is fluctuating randomly within a given range. Unbeknownst to Temp, there is an agent hidden in the room who is in control of the thermostat whose job it is to ensure that every time Temp consults the thermometer the "reading" on the thermometer corresponds to the temperature in the room. Pritchard (2012), 260.

Temp's belief is justified and safe, yet it nonetheless falls short of knowledge. The reason for this is that his belief is safely true in a deviant way. It has the wrong direction of fit with the world: the temperature of the room is adjusted to ensure that his belief is true, his belief itself is not sensitive to the temperature of the room. The virtue reliabilist is able to explain this by pointing out that Temp does not have a true belief because of his belief being the output of his epistemic abilities, rather he has a true belief because of the meddling of a third party. The proper functionalist can adopt a similar response.

A satisfactory account of NON-DEVIANCE will state what, in addition to truth and an absence of luck, is required for a belief to constitute knowledge.⁷ Does this come down to the task of providing a suitable restriction on belief forming processes? We believe that it does not. There are cases in which an agent can come to know despite their belief not being true by virtue of being an output of some reliable process of belief formation. The process restriction approach works well for cases like BRAIN and like TEMP. But it falls short of drawing the line between some genuine cases of knowledge and cases of deviant-safe-beliefs. We will therefore suggest an alternative focusing on non-deviant explanations.

In the following four sections we show how characterizing safety and non-deviance strictly in terms of methods of belief formation (as virtue reliabilists and proper functionalists do) results in misclassification of cases.

Unreliably Formed Knowledge

Consider the following case:

DOCTOR: Mark has gone to the hospital to seek a diagnosis for his apparent illness. Alex, his doctor, runs a routine test, and concludes that Mark has a brain tumor. At t_1 , Alex tells Mark "You have a brain tumor". Mark believes her. As it happens, Mark's brain has a slightly odd structure. Alex's test, although highly reliable with respect to most brains, is completely unreliable when it comes to brains structured like Mark's. Later, unbeknownst to Mark, Alex runs a routine follow up test. This second test, un-

⁷ It may be that there is no single factor capable of accounting for deviance in all cases. We may need, for example, a justification condition plus some further condition to eliminate cases in which the agent's belief is safe in a non-deviant way. However, we will be presupposing that a unified approach can and should be given. It may be possible to go one step further and provide a unified account to the anti-luck and non-deviance assumptions together. Some versions of robust virtue reliabilism attempt to do this (such as Greco (2003, 2010), and Broncano-Berrolca (forthcoming)).

like the first, is highly reliable on all brains, including those like Mark's. Alex expertly interprets the results and at t_2 she reaches two conclusions. Firstly, she finds that given the unique structure of Mark's brain the first test was unreliable. Secondly she finds out that Mark does, with utmost certainty, have a brain-tumor. If Alex discovered that Mark didn't have a brain tumor she would have informed him. Indeed, she would have been morally, socially, and institutionally obligated to do so. However, as things stand, she sees no need to tell him about the follow-up results.

Based on Alex's findings, it turns out that Mark's belief at t_1 was unreliably formed: it was luckily true, as it was based on Alex's luckily true belief. However, once the second set of tests has been performed his belief is no longer luckily true. Unbeknownst to Mark, if he did not have a brain tumor then Alex would have told him that he did not have a brain tumor. So, after t_2 , his belief could not easily have been false. Yet the basis of his belief never changed. There was no additional input; no event which provided new support for his belief. The event that changes his belief from accidentally to non-accidentally true is not the type of process that can cause new beliefs – it is not a belief forming process like testimony or visual perception.⁸

Moreover, after t_2 Mark's true belief seems to be non-deviant. His belief was justified when formed (and thereafter): he simply trusted the word of a doctor. And, in contrast with some cases where the agent's justified safe belief falls short of knowledge (cases like TEMP), Mark's belief exhibits the correct direction of fit. His belief state is, due to his social environment, sensitive to the world (not vice versa).⁹ When his belief becomes non-accidentally true it gains this status because, after the second test has been performed, everything is as it should be: There are systems of norms and practices in place to ensure that he doesn't believe falsely. These systems of norms are what brought Alex to perform a highly reliable test. And in light of those norms Alex is both equipped and disposed to ensure that Mark doesn't have a false belief about whether or not he has a brain tumor. Indeed, after t_2 Mark's belief about his brain tumor has what Goldberg (2011) calls "epistemic coverage". Given the habitual standards of

⁸ Strictly speaking Alex's discovery may still be counted as a cause of Mark's believing that he has a brain tumor at t_2 since Mark's belief at t_2 is counterfactually dependent on Alex's discovery. In the causation literature such cases are known as cases of 'double prevention': A process takes place which would cause an event A if left uninterrupted. However, external to this process a second process is taking place which would interrupt the first process and prevent A from occurring. Yet, a third event B takes place which interrupts the process which would prevent A, so despite being (potentially) spatio-temporally unconnected, and despite the fact that B does not seem in any sense to produce A, A still becomes counterfactually dependent on B. It is controversial whether causation by double prevention really counts as causation at all (see Dowe (2001) Beebe (2004), and Varzi (2007)). For example, it may be that events like B are not causes but merely part of wider causal explanations. Even those who include cases of double prevention as cases of genuine causation typically treat cases of double prevention and causation by omission to be distinct (See Hall (2004) for instance). Belief forming processes produce beliefs. The process upon which Mark's believing truly becomes counterfactually dependent in DOCTOR does not produce beliefs (for him).

⁹ Alex's insensitivity to some further norm which (if indeed applies) dictates that she should give Mark more information is irrelevant. The original testimony, the second test, and Alex's disposition to correct Mark's belief is sufficient to explain Mark's true belief, and render it non-lucky. And these factors are non-deviant. This said, we think it's clear that Alex has no such obligation.

his environment, if his belief was false he would be told so in a reliable and prompt way.¹⁰¹¹ This stands in stark contrast to the paradigm cases of strange and fleeting processes. In these cases the safety of the agent's belief was explained by the highly unusual circumstances in which they unwittingly found themselves: the kinds of circumstances we would usually be entitled to rule out.

We are assuming that a belief constitutes knowledge if it is true, non-lucky, and non-deviant. After t_2 Mark's belief that he has a brain tumor satisfies these conditions. Thus, we conclude that after t_2 Mark knows he has a brain tumor.¹² We will now explain why this supports explanation indexed safety over method indexed safety.

Safety: Methods vs Explanations

As previously mentioned, the safety principle is commonly indexed to a method of belief formation. When assessing whether an agent knows that p , safety theorists don't merely consider whether the agent could easily have falsely believed that p . Rather, they consider whether the agent could have formed a false belief that p via the same method. On this version of the safety principle Mark's belief in DOCTOR is unsafe. He formed his belief by trusting Alex's testimony at t_1 , and doing so could easily have led him astray. This fact does not change later on. Even after t_2 it is true that he formed his belief by trusting Alex's testimony, and the fact that Alex's testimony was unsafe does not somehow change between t_1 and t_2 . Thus, at t_2 Mark's belief is, for the method indexer, unsafe. Yet, it seems clear that after t_2 Mark's belief is not luckily true. So method indexed safety must be rejected.¹³¹⁴

¹⁰ Indeed, the case can be precisified in such a way that Mark could, if he so chose, reason on the basis of this fact (to which he is plausibly entitled) and thereby gain further justification for his belief. Importantly, in DOCTOR Mark doesn't reason any further after he accepts Alex's testimony. We might even assume that he is unconscious for the remainder of the case. To see that Mark's belief does not change in justificatory status, even in versions of the case where he is (or would be) aware of its coverage, consider the following analogy. Imagine I am told that $4762 + 8973 = 13735$. I come to believe this on the basis of testimony. I have the cognitive resources to verify that $4762 + 8973 = 13735$. However, if I don't utilize these resources, and I continue to believe merely on the basis of testimony, then the resources with which I *could* verify this fact don't positively impact upon my justification. This is the position Mark is in with respect to his ability to engage in epistemic coverage based reasoning. The point is that he *could* reason in such a way, and thereby gain further justification. This would not be the case if the systems which happen to support his belief were deviant.

¹¹ We are open to the possibility of cases like DOCTOR in which the belief in question does not have epistemic coverage (at least as standardly conceived), in which the truth and safety of the belief is secured in some other non-deviant way.

¹² DOCTOR, and the argument derived from it, structurally parallels Lackey's (2005) MAYOR case, whereby a normatively defeated belief comes to constitute knowledge due to the normative defeater ceasing to be present in the environment.

¹³ It might be objected that Mark's belief is based on a more extended form of trust, which includes the assumption that Alex would inform him if it was found that he didn't have a brain tumor. In this case, perhaps Mark's belief at t_2 counts as being sustained by the same method, but after t_2 his belief is safe relative to this method. It is plausible that such extended forms of trust do play an important role in our epistemic lives. There are two responses to this contention. Firstly, we can simply stipulate that Mark treats Alex's utterance as a normal case of testimony, without investing any extended trust in her. Secondly, we can modify DOCTOR in such a way that it is the actions of a third party, completely detached from the testimonial exchange, that removes the element of luck. We might suppose another doctor, Terry, takes over the case and, unbeknownst to Mark, performs the second test.

¹⁴ It is important to distinguish this problem from a related issue raised by Kelp (2014). Kelp provides a case in which an agent forms an accurate belief that it is 2.30 by observing a stopped clock which happens to be displaying '2.30' at 2.30. A technician working on the clock would correct the agent's belief if it were false. Kelp intuitively thinks that this is a case of knowledge, but points out that standard method indexed approaches to safety (along with various versions of virtue reliabilism) struggle to accommodate this intuition. This is because the agent's method, looking at a stopped clock, is arguably unreliable. (In response, Kelp suggests an alternative method in-

Thankfully there is a natural successor available: explanation indexed safety. At t_2 Mark's belief shifted from non-knowledge to knowledge. The method by which his belief was formed does not change. He does not gain any new evidence, there is no input from any belief forming process. What does change is the explanation for his continued belief. Before t_2 his belief is explained entirely by the method by which it was formed plus the relevant background conditions. After t_2 these factors no longer suffice to explain his belief. Given that the second test has been performed, and given Alex's dispositions, at this point the outcome of the second test becomes explanatorily relevant to his continued belief. Thus, his belief becomes counterfactually dependent on its truth.

By indexing safety to the explanation for the agent's belief at a time (or to "time indexed explanations") we are able to secure the result that Mark's belief is safe at t_2 and unsafe at t_1 . At t_1 , given the explanation for his belief, Mark could easily have falsely believed that he had a brain tumor. After t_2 , given the explanation for his belief, he could not easily have falsely believed that he had a brain tumor.

In addition to making the correct prediction about DOCTOR, explanation indexed safety is able to make the same predictions as method indexed safety in cases where the agent's belief is explained by the process by which it was formed. If we restrict our attention to such cases, then explanation indexed safety comes out as equivalent to method indexed safety. This explains why method indexing seems so natural, despite being false.

The Problem of Strange and Fleeting Explanations

In order to address the problem of strange and fleeting processes, virtue reliabilists and proper functionalists place a restriction on admissible belief forming processes: Beliefs are non-deviant when they are safe by virtue of being formed by such processes.

We have argued that safety should be indexed not to methods but rather to time indexed explanations. We believe that a satisfactory account of non-deviance should explain what it is for a time indexed explanation to be non-deviant. As we argue, failing to account for non-deviant explanations will prevent us from drawing the line between some cases of genuine knowledge (like DOCTOR) and cases of deviant-safe-beliefs. To wit: a problem parallel to Greco's problem of strange and fleeting processes arises at the level of explanations. There are cases in which the agent fails to gain knowledge because the explanation for their belief is deviant. Consider the following case:

MAD SCIENTIST: Samantha is receiving brain surgery. What she doesn't know is that her doctor is an all knowing mad scientist with the ability to remove beliefs from his patients. During the surgery the doctor decides to remove all of Samantha's false

dexed approach according to which the technician's monitoring is built into the method). Note, however, that whenever we assess a belief for safety, apart from the method in question, we must hold certain background conditions fixed. The standard method indexer can therefore simply claim that the technician's presence and dispositions are part of the epistemic environment within which the belief was formed, and must thus be held fixed along with the method. Indeed, this is exactly the conclusion Goldberg (2005) draws when considering a parallel case involving unsafe testimony. Either way, Kelp's cases are not designed to threaten method indexing. This stands in contrast to DOCTOR. In DOCTOR Mark's belief is unsafe when formed. There is no background environmental condition which would render his belief safe (if held fixed when assessing his belief for safety). Nor can the later test be built into the method by which Mark forms his original belief.

beliefs. He doesn't remove Samantha's gettiered belief that p , because p is true. However, if p was false then he would have removed it.

This case is structurally parallel to DOCTOR. Until the surgery Samantha's belief that p is explained in terms of the process by which it was formed (which we can assume was non-deviant). After the surgery these factors are no longer sufficient to explain her continued belief, for her belief becomes counterfactually dependent on its truth. It is, thus, safe. Yet Samantha, unlike Mark, does not gain knowledge. The reason for this is that unlike the explanation for Mark's belief at t_2 , the explanation for her continued belief in p , the mad scientist's invasive meddling, does not seem to be of the right kind to support knowledge.

This suggests that a unified account of non-deviance will have to explain not just what it is for methods or belief forming processes to be deviant, but also (more generally) what it is for time indexed explanations to be deviant.¹⁵ Since time indexed explanations will often be given in terms belief forming processes, such a theory must explain what it is for a belief forming process to be deviant. It must, for example, explain what is wrong with the processes by which Sandy and Temp form the beliefs discussed in BRAIN and TEMP (respectively). However, any such account must generalize beyond belief forming processes, and apply at the level of explanations. This is necessary for distinguishing cases like MAD SCIENTIST from cases like DOCTOR. In the next section, we will briefly explain why the dominant approaches to non-deviance; virtue reliabilism and proper functionalism, struggle to do this. We will close by introducing an alternative.

Non-Deviant Explanations

The dominant approach to the problem of strange and fleeting processes is virtue reliabilism. Virtue reliabilists hold that one has knowledge only if one has a true belief because of the exercise of one's cognitive abilities. There are different ways of spelling this out depending on the way one reads the 'because',^{16,17} but this need not concern us here. It is hard to see how any such view can be generalized to the level of explanations and thereby capture the fact that the explanation for Mark's belief in DOCTOR, but not Samantha's belief in MAD SCIENTIST, is non-deviant. Given DOCTOR, it seems that an agent's cognitive abilities need not *fully* explain their true belief in order for it to constitute knowledge. But obviously, the simple fact that an agent's cognitive abilities play *some* role in the explanation of their true belief is not sufficient to rule out deviance. After all, Temp's belief (TEMP) and Samantha's belief (MAD-SCIENTIST) are both partly explained by the cognitive abilities they manifested when they formed their beliefs. The explanatory role played by *their* cognitive abilities is not in any apparent way less dominant than explanatory role played by Mark's.¹⁸ It is thus unclear how the virtue reliabilist can explain why Mark's belief, but not

¹⁵ As with TEMP and BRAIN one might opt for a disunified approach, treating MAD SCIENTIST separately. We will not consider such approaches here. We assume that a unified approach can and should be given.

¹⁶ Greco (2010) and Pritchard (2012) adopt a causal explanatory account of the 'because'. They require that the agent's exercise of their causal abilities be a salient explanatory factor in securing their true belief. Turri (2011), Sosa (2011, 2015), and Kelp (2013) adopt a manifestation reading of the 'because' according to which the agent's true belief must manifest their cognitive abilities.

¹⁷ There is also an important division amongst virtue reliabilists between those who, like Pritchard (2012) and Kelp (2013) combine the restriction on admissible belief forming processes with a safety principle, and those who, like Greco (2010) and Broncano-Berrocal (forthcoming), hold that virtue reliabilism is able to rule out luck without appeal to an explicit anti-luck condition.

¹⁸ Kelp (2013) presses a similar objection against Pritchard's anti-luck virtue epistemology, holding that the competencies involved in the acceptance of testimony are explanatorily on a par with the competencies involved in

Samantha's or Temp's, is non-deviant.¹⁹ Since replacing 'cognitive abilities' with 'properly functioning belief forming process' yields similar results, proper functionalism succumbs to the same problem.

The first natural response to this problem, for the virtue reliabilist at least, is to hold that knowledge needn't be an individual achievement, but may instead be a group achievement. This is precisely the move that Sosa (2007) makes to capture the sense in which testimonial knowledge is an achievement. He draws an analogy with the credit a quarterback might deserve for a touchdown pass in an American Football game:

A quarterback may throw a touchdown pass, for example, thus exercising a competence. But this individual competence is only one part of a broader competence, seated in the whole offensive team, that more fully explains the successful touchdown pass, the apt performance of that quarterback. The pass receiver's competence may be crucial, for example, along with the individual competences of the offensive linemen, and so on. Sosa (2007), 94.

It might be thought that Mark's belief is not an individual achievement, but a group achievement. He deserves some credit for his belief, as he exercised some cognitive abilities in forming his belief. But the credit is shared with Alex. This allows the virtue reliabilist to make the correct prediction about DOCTOR. However, this success comes at the cost of making the wrong prediction about MAD SCIENTIST.

In MAD SCIENTIST, like in DOCTOR, there is a sense in which Samantha's safe true belief constitutes a group achievement. Like Mark, Samantha deserves minimal credit for her true belief, as she exercised her cognitive abilities in forming her belief. However, after t_2 the lion's share of the credit for her true belief (and all the credit for its safety) should go to the mad scientist. It was the mad scientist's exercise of their cognitive abilities which explains why Samantha's belief is safe.

In order to secure the right prediction, Sosa needs a non-arbitrary way to restrict the social groups who's collective achievements are able to support knowledge. That is, he needs a way of saying that Samantha and the mad scientist (or Temp and the temperature manipulator for that matter) do not count as groups in the relevant sense, whilst Mark and Alex do. We do not see how such a restriction could be given.²⁰ This is not quite a knock down objection to the

cases like TEMP. Following Sosa (2007, 2010), Kelp holds that manifestation of an ability requires that various internal and external preconditions be satisfied. This allows Kelp to hold that testimonial knowledge manifests one's competence, whereas Temp's true belief (due to the unfavorable environmental conditions) does not. Kelp's response does not carry over to DOCTOR, as the explanation for Mark's true belief after t_2 does not involve the manifestation of his cognitive abilities.

¹⁹ Note that the problem is not that virtue reliabilists are unable to account for the impact of the agent's epistemic environment on the epistemic status of their belief. This problem, as opposed to the problem of strange and fleeting explanations, is a problem which some virtue reliabilists, those that incorporate a separate anti-luck condition do not face (see Pritchard (2015) and Pritchard and Kallestrup (2014, 2016)).

²⁰ As Sosa himself points out, groups must be construed very loosely for the social achievement view to be capable of capturing many cases of testimonial knowledge. For example, a historic letter writer and a modern historian hardly form a cohesive social group, nor do a speaker and an eavesdropper. But in both cases the latter can gain testimonial knowledge from the former. Indeed, as we noted previously, DOCTOR can be modified such that Terry, a doctor who Mark has never met nor interacted with, performs the second test. Mark's belief would still become non-accidentally true in this scenario. It is hard to see why Mark and Terry (or, for that matter, the technician and belief former in Kelp's stopped clock case), who may never have even interacted, would form a more cohesive social group than Samantha and the surgeon.

virtue reliabilist. Further research may reveal a viable notion of group achievement capable of delineating the cases in the correct way. However, we think this is a big enough problem to motivate the search for an alternative.

Another response takes its cue from Kelp's (2014) extended-epistemology, according to which belief forming processes may extend beyond the agent's skin. Drawing on Clark and Chalmers (1999), Kelp includes in the agent's belief forming process any process such that if it occurred inside the agent's head we would have no problem counting it as part of the belief forming process (ibid. 244-5). An agent's true belief can then be thought to manifest the competence of an extended cognitive system. This response could be extended to DOCTOR if we consider Mark and Alex to be part of such an extended cognitive system. We could say that Mark's believing truly after t_2 manifests a competence of this extended cognitive system.

Kelp's approach bears clear similarities to Sosa's. Both approaches move from treating knowledge as an achievement of the individual knower to an achievement of (or a manifestation of the competence of) some more extended system. Unlike Sosa, however, Kelp provides us with a principled way of ruling out certain deviant processes. Any process which we would not count as part of the belief formation (or retention) process if internalized should not be counted. This arguably enables Kelp to accommodate cases such as TEMP, as the temperature manipulator's meddling would not, if internalized constitute part of Temp's belief forming process. Unfortunately, Kelp's approach, like Sosa's, is unable to secure the correct result with respect to MAD SCIENTIST. In MAD SCIENTIST the scientist sorted through Samantha's brain checking each belief for truth, and keeping any beliefs which were true. If this process was internalized, and Samantha was doing the sorting, we would have no problem saying that Samantha's belief is retained on the basis of this process. Thus, Samantha's true belief would manifest the competence of this extended system. So like other virtue reliabilists, Kelp is going to have trouble in drawing the line between genuine knowledge and deviant safe beliefs.

Virtue reliabilism and proper functionalism fail to accommodate the shift to explanation based safety and fail to account for non-deviance at the level of explanations. It should be clear how this problem generalizes to other views which emphasize the production of beliefs, rather than their explanations more generally. Any such approach will be unable to provide a unified approach to non-deviance, capable of capturing cases of deviant explanation more generally. We will close by considering an alternative. We do not claim that this is the only viable approach to non-deviance. However, it functions to illustrate the shape a successful approach to knowledge must take in light of the considerations presented above.

Explanation and Normality

We need a new approach to non-deviance. We need a normative notion like those employed by the virtue reliabilists and the proper functionalists, but which applies at the level of explanations. In order to achieve this we propose bringing into the foreground a notion which already plays an essential background role in many current versions of reliabilism, including

virtue reliabilism and proper functionalism: the notion of normality.²¹ We will close this paper by providing the beginnings of a normality based approach to knowledge.²²

There are at least two notions of normality in our pretheoretic conceptual repertoire. There is a statistical notion on which normality reduces to statistical prevalence. However, as Nickel (2008, 2016) and Smith (2010) observe, there also seems to be a non-statistical notion of normality, a notion according to which the normal is explanatorily privileged over the abnormal. This is, plausibly, the notion of normality we evoke when we make certain generic generalizations such as “Turtles live to an old age” or “Tigers have stripes” (Asher and Pelletier (1997, 2012), and Nickel (2008, 2016) provide normality based approaches to generics).^{23,24} When we make such generalizations we are not simply stating that tigers usually have stripes, or that turtles usually live to an old age. If this were what we were doing then we would be speaking falsely when we say that turtles live to an old age, since the majority of turtles die in their first moments of life. Rather, when we make such generalizations we are conveying that the attributed property is normal for the kind in question; it is characteristic. Turtles characteristically live to an old age. Likewise, tigers characteristically have stripes. If a tiger fails to have stripes then this cries out for explanation.²⁵

For an exemplar of a kind to be normal in the non-statistical sense it must possess the characteristic properties of the kind it exemplifies. For example, for a raven to be normal it must be able to fly, it must be black, it must have a beak etc. However, this is not sufficient for normality. In order to be normal an object must possess these characteristic properties in a normal way. An albino raven which is painted black is not normal, despite having the characteristic property of being black. Being painted is not a normal way for ravens to be black. The characteristic way for a raven to be black is as a result of genetic inheritance. If a raven is black despite not having inherited its blackness then its blackness cries out for explanation (we take this example from Nickel (2016)). This gives us the key ingredients we

²¹ Leplin (2007, 2009) and Graham (2016) provide reliabilist accounts of justification according to which a belief is justified iff the belief is caused by a process which reliably causes true beliefs in normal circumstances. Sosa (2007) characterizes competencies as belief forming processes which reliably produce true beliefs in normal circumstances. Greco (2008) appeals to the notion of normality in characterizing the conditions in which we assign competencies explanatory salience. In Gettier cases, he tells us, the default salience of the agent’s cognitive competencies in explaining their true belief is trumped by some abnormal intervening factor.

²² We only aim to introduce the normality view as an alternative to virtue reliabilism and proper functionalism, not provide a full explication and defense. The latter would require far more space than we have available.

²³ Nickel also provides a worked out account of normality. For Nickel, an exemplar of a kind is normal if it has the characteristic properties of its kind, and has them in a characteristic way, viz. in virtue of the mechanisms evoked by the explanatory strategies relative to which the property counts as characteristic. More strictly: a property *F* is characteristic of a kind *K* relative to a set of explanatory strategies *S* if the presence of *F*ness among the *K*s can be explained by *S*. For example, “being stripy” is present among the tigers, and this can be explained in terms of systems of biological inheritance. Importantly: we employ different types of explanatory strategy for different kinds. For example, we judge that it is normal for ravens to be black because the blackness of ravens is explained in terms of inheritance. It is normal for alarm clocks to ring at set times because of the way in which they are designed. There is also a degree of context sensitivity here, as different explanatory strategies (e.g. design, inheritance etc.) will be relevant in different contexts. This bares similarities to Smith’s (2010) approach according to which the normal constitutes an explanatory default, deviations from which cry out for explanation. If *a*’s *G*ness cannot be explained by the salient explanatory strategies at play in the context, then *a*’s *G*ness seemingly cries out for explanation.

²⁴ We do not claim that generics always express claims about normality, or that an adequate semantics for generics must make reference to the notion of normality. We merely claim that certain uses of generics, those which seem to characterize their targets, express claims about normality.

²⁵ These claims might also be expressed by ‘oughts’ of normality. There is a sense in which tigers ought to have stripes and turtles ought to live to an old age. See McGrath (2005) for an account of non-statistical normality in terms of such ‘oughts’.

require for our account of knowledge. We propose that knowledge is (non-statistically) normal belief.²⁶ Or, to be more exact, that knowledge is characteristically safe true belief. Truth and safety are characteristic features of belief. Beliefs are often false, and even when true they are sometimes unsafe. However, when things are as they should be beliefs are true and safe.²⁷ In cases like MAD SCIENTIST, and TEMP the beliefs have the characteristic properties of truth and safety. Yet, like the painted albino raven, they do not have their characteristic properties in characteristic ways. Thus, TEMP and MAD SCIENTIST are not cases of characteristically safe belief. However, in DOCTOR everything is as it should be; Mark's safe true belief is explained in terms of the systems of norms and practices which are in place to ensure that he doesn't believe falsely. His belief not only has the characteristic properties of truth and safety. It has these properties in a normal way. Thus, the view that knowledge is characteristically safe true belief provides a straightforward solution to the problem of strange and fleeting explanations. The problem with explanations of this sort is precisely that are not normal; they are strange.

We have left open the question of which explanations for safe true belief are normal. Just as it is the task of the biologist to uncover the normal mechanisms which give rise to ravens' blackness, it is the task of epistemologists to examine the normal mechanisms by which we come to have safe true beliefs. This involves studying the mechanisms by which we acquire true beliefs, mechanisms such as perception, memory, and testimony, and explaining how these mechanisms give rise to true non-lucky beliefs. By this point there exists, for any given mechanism, a plethora of competing explanations for its ability to produce non-lucky true belief.²⁸ However, we believe that a complete epistemology will look not only at the mechanisms by which we normally acquire knowledge, but also the wider systems of norms (both institutional and social) which support and explain the safety of our beliefs.

Bibliography

- Asher, N., Pelletier, F. 2012. More Truths about Generic Truth. In Mari, Beyssade, and Prete (eds) *Generativity*. Oxford University Press. 312-333.
- Ball, B. 2013. Knowledge is normal belief. *Analysis*. 73 (1). 69-76.
- Beebe, H. 2004. Causing and Nothingness. In Collins, Hall, and Paul (eds) *Causation and Counterfactuals*. Cambridge, MA: MIT Press. 291-308.
- Bogardus, T. (2014). "Knowledge Under Threat," . *Philosophy and Phenomenological Research*. 88, pp. 289–313.
- Broncano-Berrocal, F. Forthcoming. Purifying Impure Virtue Epistemology. *Philosophical Studies*.
- Clark, A. Chalmers, D. 1998. The Extended Mind. *Analysis*. 58. 7-19.
- Comesaña, J. (2005). "Unsafe Knowledge," *Synthese*. 146. 395–404.
- Fricker, E. 2016. Unreliable Testimony. In Kornblith and McLaughlin (eds) *Alvin Goldman and his Critics*. Blackwell.

²⁶ Ball (2013) also proposes that knowledge is normal belief. However, Ball locates this hypothesis within a more traditional proper functionalist framework.

²⁷ It is important when we assess whether a particular belief constitutes knowledge that we consider it qua belief, not qua belief of some more restricted kind. Safety and truth is characteristic of belief, but not of belief formed via the gamblers fallacy.

²⁸ Indeed, it still seems probable that there are important general claims which can be made about all belief forming methods capable of yielding knowledge. For example, it is consistent with what we have argued here that cognitive ability must always be a contributing factor, in at least some very weak sense. Our point is that such an insight, if it is correct, cannot do the fundamental work many epistemologists assign to it.

- Dowe, P. 2001. A counterfactual theory of prevention and 'causation' by 'omission'. *Australasian Journal of Philosophy*. 79 (2). 216-226.
- Gettier, E. 1963. Is Knowledge Justified True Belief? *Analysis*. 23 (6). 121-123.
- Goldberg, S. 2005. Testimonial Knowledge from Unsafe Testimony. *Analysis* 65. 302-311.
- Goldberg, S. 2011. If That Were true I Would Have Heard it by Now. From Goldman and Whitcomb (eds) *Social Epistemology: Essential Readings*. Oxford University Press. 92-108.
- Goldberg, S. 2015. Epistemic Entitlement and Luck. *Philosophy and Phenomenological Research*. 91 (2). 273-302.
- Graham, P. 2016. Against Actual-World Reliabilism: Epistemically Correct Procedures and Reliably True Outcomes, In Fernandez (ed) *Performance Epistemology: Foundations and Applications*. Oxford University Press. 83-106.
- Graham, P. Forthcoming. Normal Circumstances Reliabilism. *Philosophical Topics*.
- Greco, J. 1999. Agent Reliabilism. *Philosophical Perspectives*. 13. 273-296.
- Greco, J. 2003. Knowledge as Credit for True Belief. In DePaul and Zagzebski (eds) *Intellectual Virtue: Perspectives from Ethics and Epistemology*. Oxford University Press.
- Greco, J. 2007. The Nature of Ability and the Purpose of Knowledge. *Philosophical Issues*. 17. 57-69.
- Greco, J. 2010. *Achieving Knowledge*, Cambridge: Cambridge University Press.
- Hall, N. 2004. Two Concepts of Causation. in Collins, Hall and Paul (eds) *Causation and Counterfactuals*. MIT Press. 225-277.
- Kallestrup, J., Pritchard, D. 2014. Virtue Epistemology and Epistemic Twin Earth. *European Journal of Philosophy*. 22. 335-357.
- Kallestrup, J., Pritchard, D. 2016. Dispositional Robust Virtue Epistemology versus Anti-Luck Virtue Epistemology. From Fernandez (ed) *Performance Epistemology: Foundations and Applications*. Oxford University Press. 31-47.
- Kelp, C. 2013. Knowledge: The Safe Apt View. *Australasian Journal of Philosophy*. 91 (2). 265-278.
- Kelp, C. 2014. Epistemology Extended. *Philosophical Issues*. 24. 230-252.
- Lackey, J. 2005. Memory as a Generative Epistemic Source. *Philosophy and Phenomenological Research*. 70 (3). 636-658.
- Lackey, J. 2006. Pritchard's Epistemic Luck. *The Philosophical Quarterly*. 56. 284-289.
- Lackey, J. 2007. Why We Don't Deserve Credit for Everything We Know. *Synthese* 158. 345-361.
- Lackey, J. 2008. What Luck is Not. *Australasian Journal of Philosophy*. 86 (2). 255-267.
- Lackey, J. 2009. Knowledge and Credit. *Philosophical Studies* 142. 27-42.
- Leplin, J. 2007. In Defense of Reliabilism. *Philosophical Studies*. 134 (1). 31-42.
- Leplin, J. 2009. *A Theory of Epistemic Justification*. Dordrecht: Springer.
- Neta, R. and Rohrbaugh, G. (2004). Luminosity and the Safety of Knowledge. *Pacific Philosophical Quarterly*. 85, .pp. 396-406.
- Nickel, B. 2008. Generics and the Ways of Normality. *Linguistics and Philosophy*. 31 (6). 629-648.
- Nickel, B. 2016. *Between Logic and the World: An Integrated Theory of Generics*. Oxford University Press.
- Nozick, R. 1981. *Philosophical Explanations*. Harvard University Press.
- Pelletier, F. Asher N. 1997. Generics and Defaults. In van Benthem and A. ter Meulen (eds.) *Handbook of Logic and Language*. MIT Press.. 1125-1179.
- Plantinga, A. 1993. *Warrant: The Current Debate*. Oxford University Press.
- Pritchard, D. 2005. *Epistemic Luck*. Oxford: Clarendon Press.
- Pritchard, D. 2012. Anti-Luck Virtue Epistemology. *The Journal of Philosophy*. 109. 247-79
- Pritchard, D. 2014. Epistemic Dependence. *Philosophical Perspectives*. 29. 305-324.

- Smith, M. 2010. What else justification could be. *Noûs*. 44 (1). 10-31.
- Sosa, E. 2007. *A Virtue Epistemology*. Oxford: Clarendon Press.
- Sosa, E. 2010. How Competence Matters in Epistemology. *Philosophical Perspectives*. 24. 465-476.
- Sosa, E. 2011. *Knowing Full Well*. Princeton University Press.
- Sosa, E. 2015. *Judgement and Agency*. Oxford University press.
- Turri, J. 2011. Manifest Failure: The Gettier problem Solved. *Philosopher's Imprint*. 11. 1-11.
- Turri, J. 2013. Unreliable Knowledge. *Philosophy and Phenomenological Research*. 90 (3). 529-545.
- Varzi, A. 2007. Omission and Causal Explanation. In Castellani and Quitterer (eds) *Agency and Causation in the Human Sciences*. Pader born: Mentis Verlag. 155-167.