Reliabilism Defended

Jeffrey Tolly

Reliabilism about knowledge states that a belief-forming process generates knowledge only if its likelihood of generating true belief *exceeds* 50%. Despite the prominence of reliabilism today, there are very few if any explicit arguments for reliabilism in the literature. In this essay, I address this lacuna by formulating a new independent argument for reliabilism. As I explain, reliabilism can be derived from certain key knowledge-closure principles. Furthermore, I show how this argument can withstand John Turri's two recent objections to reliabilism: the argument from explanatory inference and the argument from achievements.

Knowledge, Reliabilism, The Generality Problem, Closure, Contrastivism

[total word count including references and notes: 10,003]

1. Introduction

According to reliabilism about knowledge (henceforth, "reliabilism"), a belief b counts as

knowledge only if b is produced by a process, disposition, or ability that tends to generate

more true beliefs than false beliefs, i.e., has a truth-ratio over 50%. Reliabilism is a dominant

viewpoint in contemporary (post-Gettier) epistemology.[1] But not everyone is so convinced.

---

[1] For example, Alvin Goldman (1979), Duncan Pritchard (2009:415), and Ernest Sosa (2007:29) all voice their support for reliabilism about knowledge in one way or another. John Turri aptly captures the near consensus on reliabilism as follows: "Adapting a Quinean coinage, it's not unfair to label knowledge reliabilism the central dogma of contemporary epistemology" (2016: 190).

For starters, John Turri points out, "[t]he literature contains surprisingly little explicit argumentation for knowledge reliabilism…It is often just asserted, without elaboration, that knowledge requires reliability" (2016: 188). On top of this observation, Turri has recently produced two arguments for the possibility of unreliably produced knowledge, or "unreliable knowledge" for short: the argument from explanatory inference and the argument from achievements. This constitutes a serious challenge to the reliabilist orthodoxy of today—a challenge I endeavor to meet in this essay.

In §2, I address Turri's initial worry by offering a new argument for reliabilism. As I'll show, the fact that knowledge is closed under certain entailment relations allows us to derive reliability conditions on both knowledge-that (i.e., binary knowledge) *and* knowledge-wh (i.e., knowledge of answers to open-ended questions). This is an important result, as Turri takes certain instances of knowledge-wh to constitute the clearest counterexamples to reliabilism. Moreover, insights from this closure argument for reliabilism highlight critical problems with Turri's two objections to reliabilism, which I address in the second half of the paper. In §3, I respond to the argument from explanatory inference, showing how it depends on adopting an independently implausible solution to the generality problem for reliabilism. Finally, in §4, I reply to the argument from achievements. I explain how Turri's observations about other unreliably produced achievements fail to support the possibility of unreliable knowledge. I conclude that neither of Turri's arguments rebut or undercut the closure-based argument for reliabilism.

2. An Argument for Reliabilism

2.1   Closure and Reliabilism: a First Pass

In developing an argument for reliabilism, our point of departure—ironically enough—is nicely articulated by John Turri.

> [W]e can glimpse one reason why truth-conducive reliabilism might seem unavoidable. Suppose that when we're considering whether someone knows Q, we think, "In order for her to know Q, she must have an ability to get at the truth of the matter. And if she has such an ability, then she gets the truth at a rate better than chance. Moreover, here chance is 50/50 because there are only two options: *either Q is true, or it isn't.* So knowledge requires truth conducive reliability." (2015: 542, emphasis mine)[2]

Here, I wholeheartedly agree; this line of reasoning does make reliabilism seem unavoidable. Importantly, we can take the central idea of this passage and incorporate it into an argument for reliabilism, which I formulate in this subsection. Then, in §2.2, I'll discuss an important limitation of this argument.

An argument for reliabilism emerges once we consider how knowledge is closed under at least some entailment relations. In particular, knowing p puts the subject in a position to know certain things about the process used for obtaining her knowledge that p, and this feature of knowledge helps illuminate why a 50% chance of delivering truth is *too* chancy for knowledge production. Suppose someone both knows p and knows she used process P to form her belief that p. Presumably, such an agent is also in a position to know that process P didn't generate a false belief *on the particular occasion* when she used it to form her belief that p. After all, conjunctive statements of the following form seem patently absurd: "Claim p is true and I can tell from using process P, but who's to say whether P delivered true or false output when I used it to arrive at my belief that p." The following closure principle nicely accounts for the absurdity of such conjunctions:

---

[2] Here, a process is "truth conducive" only if its relevant truth ratio exceeds 50%.

Process Closure (PC)

Necessarily, where q is a proposition of the form *S used process P to form her belief p,* r is a proposition of the form *P didn't produce a false belief when S used P to form her belief p,* and where S knows both p and q,

if S competently deduces r from p and q, then S knows r.

PC is the first premise in a knowledge-closure argument for reliabilism, which runs as follows:

R1.   PC
R2.   If unreliable knowledge is possible, then PC is false.
R3.   Therefore, unreliable knowledge is impossible.

Here's the basic idea behind R2: if unreliably produced knowledge is possible, then possibly there are exceptions to PC, i.e., then it's possible that there are violations of the necessary entailment relation described in PC.

R2 is supported, in part, by Turri's ideas from the above quotation. Let's assume for the sake of argument that unreliable knowledge *is* possible, and that an agent S_a could come to know some claim p_a through the use of a process P_a that's 50% reliable (or less). If unreliable knowledge is possible, there's no in-principle reason for denying the possibility that S_a could also come to know claim q_a (that process P_a produced her belief that p_a). Furthermore, there's no reason for denying that S_a could also competently perform the deduction described in PC. But then, given PC, S_a would then *know* claim r_a (that her process P_a didn't generate a false judgment when it produced her belief that p_a).

Yet, it's implausible that this deduction delivers knowledge of r_a. In performing this deduction, S_a would, in essence, judge that a process on one particular occasion didn't deliver a false output simply on the basis of the very output generated on that occasion. But intuitively, if a process delivers truth only 50% of the time, one couldn't come to know that the process didn't deliver a false output on a given occasion if one's basis or premise is just whatever output was delivered on that occasion. Recalling Turri's idea above, were any such

deduction to deliver a true belief, it would do so only by chance. For instance, even if P_a in fact generated a true output on the occasion when it produced S_a's belief that p_a, there was an equal chance that it would have produced a false output. This being the case, it's doubtful that S_a can simply rely on P_a by using output p_a as a premise and thereby come to *know* that P_a didn't produce a false output on that occasion.

One way to avoid this result is to deny PC. But PC is highly plausible, and the cost of denying it is prohibitively high. The reasonable option we're left with is to deny our initial assumption that unreliable knowledge is possible.

2.2   Knowledge-wh and defending unreliable knowledge

Initially, R1-R3 might look like a compelling argument for a general reliability condition on knowledge. However, there's reason to doubt that this argument could establish such a broad conclusion. Turri expresses this concern in the subsequent discussion of his envisaged defense of reliabilism, which I quoted above.

> But we shouldn't accept this reasoning [from earlier on 542]. It takes too narrow a view of the potential options, focusing myopically on Q's truth or lack thereof. Sometimes we're faced with the binary question 'is Q true?'. But often we're faced with open-ended questions, such as 'what condition is causing his symptoms?', 'when will it happen?', 'who committed the crime?', or 'why is the honeybee population declining?' (see Schaffer 2007[a]). It's no accident that one of my two main arguments against truth-conducive reliabilism featured explanatory reasoning: explanatory reasoning is our main tool for answering such open-ended questions. *It is precisely these cases that the binary model poorly fits.* (542, emphasis mine)

Here, two clarifying points are in order. First, earlier in his paper, Turri identifies a particular kind of explanatory belief-forming process that, by his lights, *can* produce unreliable knowledge. He draws our attention to cases where the subject is aware of a large class of candidate explanations for some body of evidence D. The subject goes on to believe that one of these explanations H is correct given his awareness that H is far more likely than any of the

other competitor explanations taken individually. However, the subject is also aware that, conditional on D, H is only 50% likely or less (537). For short, let's call any process that fits this general description an *unlikely best alternative process* (UBA process). Turri states, "in such a case, it's reasonable to accept that H explains D. And if it's true that H explains D, it seems that you could thereby know that H explains D"—this despite the fact that UBA processes are clearly unreliable (537). After all, H is "by far the best explanation" of the subject's evidence (537).

Secondly, as Turri cites, the view that knowing is—at least in some cases—fundamentally a matter of *answering a relevant question* squares nicely with Jonathan Schaffer's contrastivist account of knowledge. According to Schaffer, there's an important distinction between knowledge-wh—i.e., "knowledge-*who, what, when, where,* and *why,"* and knowledge-that, i.e., binary knowledge (2007a: 383). On Schaffer's account, knowledge-wh amounts to knowing the answer to a particular question Q (385). For instance, by glancing out the window, one might know *a pigeon* is the answer to the question, "Is that a pigeon in the tree, or a dog?" but not know *a pigeon* is the answer to the question, "Is that a pigeon in the tree, or a dove?" For Schaffer, question utterances "denote"—in a contextually dependent way—a set of "relevant alternative" answers (388). This being the case, knowing *p is the answer to Q* requires one to "eliminate" (or, be in an epistemic position to eliminate) the logical space corresponding to all of Q's non-p alternatives (2007b: 238).

Notably, Schaffer argues that knowledge-wh cannot be reduced to knowledge-that.[3] For the purposes of this paper, I'll assume that his argument is successful. If knowledge-wh cannot be reduced to knowledge-that, we shouldn't think that a reliability condition on knowledge-that straightforwardly entails a reliability condition on knowledge-wh. In addition, Turri

---

[3] See Schaffer (2007a: 386-9) for a full discussion of this argument.

suggests that features of UBA processes provide us with independent grounds for doubting that an argument like R1-R3 could ever establish a reliability condition on knowledge-wh. To reconstruct Turri's reasoning, it's instructive to consider a specific example of a UBA process.

COLD

> Tim is a biologist studying infectious diseases. The local hospital allows Tim to shadow doctors and patients. One week, the first patient that Tim observes manifests a set of mild symptoms D_s. Tim knows that, across a large and representative data set N cataloging cases of D_s in his local region, there are 101 distinct (*and* mutually exclusive) possible conditions that can cause D_s. According to N, 50% of D_s cases are caused by the common cold; call this hypothesis H_1. The other possible causes, H_2-H_{101}, are much less common—each (individually) only occurs in .5% of all cases of D_s in his region, and Tim knows this to be the case. Based on observing the patient to have D_s and his understanding of the statistical data N, Tim comes to believe that this patient's symptoms are caused by the cold. For short, let "P_T" denote this belief-forming process/method. In this case, P_T delivers the correct diagnosis. [4]

In COLD, Tim answers the question Q_C, "Which condition causes the patient's D_s symptoms—H_1 or H_2 or…H_{101}?" In this context, Q_C denotes the set of alternatives H_1-H_{101}, which in turn denotes a partition of logical space corresponding to this set of alternatives. If Turri's comments on UBA processes are correct, then process P_T allows Tim to eliminate H_2-H_{101} given that H_1 is far more likely than each of these other alternatives taken individually. As a result, P_T enables Tim to know H_1 is the answer to Q_C, thereby giving him knowledge-wh.

With the details of a UBA process in clear view, we can more precisely capture Turri's criticism of R1-R3 in the following way. It looks as if premise R2 rests on the assumption that the *likelihood of generating true belief rather than false belief* is the salient feature that determines whether a process generates knowledge. While this feature might be a salient determining

---

[4] See Roeber (2020: 861) for a discussion of a case very similar to COLD as part of his discussion of *improbabilism,* the thesis that it is possible to know p despite having a credence in p that is under .5 (839, 860).

factor for producing knowledge-that, it's not at all clear that it's a salient determining factor for producing knowledge-wh. Instead, perhaps the process's salient knowledge-wh-determining feature is something like *proficiency in selecting the correct answer amongst the set of alternatives*. Even though P_T is just as likely to produce false beliefs as true beliefs, P_T does seem—in some sense—proficient in selecting the correct answer amongst the set of alternatives since P_T selects H_1, and H_1 is way more likely than each of the other alternatives taken individually. This being the case, we should think that premise R2—as stated—is unmotivated. At best, we're only justified in believing a narrower variant of R2 that only applies to knowledge-that.

Here, reliabilists might concede that reliabilism about knowledge-wh lacks support, while maintaining that R1-R3 still provides adequate grounds for reliabilism about knowledge-that. However, such a concession is unnecessary. In what follows, I'll give an alternative closure argument for reliabilism about knowledge-wh. There, I'll explain why UBA processes cannot generate knowledge-wh. Importantly, this alternative argument in no way presupposes that a process's *likelihood of delivering truth over falsehood* is the salient knowledge-determining feature for that process.

2.3   Closure and reliabilism about knowledge-wh

To begin, recall Turri's main explanation for why agents like Tim can have knowledge-wh in cases like COLD: given Tim's use of P_T, the patient's having the cold is way more likely than each of the other alternatives *taken individually*. This idea suggests that there's an important epistemic difference between cases like COLD, where the relevant question has numerous alternatives, and cases where the relevant question just has two alternatives—cases like the following:

T-CELL
>All of the details are the same as COLD except for these key changes. According to data set N, 50% of patients who manifest symptoms D_s have increased T-cell counts, and the other 50% have normal T-cell counts. This is the only information contained in N. On a given instance of observing a patient to have D_s, Tim comes to believe the following claim on the basis of this observation and his background knowledge of N: *this patient has an increased T-cell count rather than a normal T-cell count.* Moreover, Tim's judgment in this particular case is correct.

While there's some intuitive pull towards ascribing knowledge-wh in COLD, it's highly doubtful that Tim's statistical belief-forming process in T-CELL allows him to know *this patient has an increased T-cell count.* Upon reflection, the key knowledge-undermining factor in T-CELL readily suggests itself: relative to Tim's belief-forming process, the selected answer and the (lone) rejected alternative are *equiprobable.* For short, we can characterize the kind of process used in T-CELL as follows:

Unlikely Two-Alternative Process (UT-process)
>Belief-forming process P is a UT-process just if P is used to answer a question Q with two alternatives and each alternative is equally likely relative to P.

In T-CELL, it seems that if Tim's UT-process were to have delivered a correct answer on this given instance, it would have done so only by sheer chance (as Turri might put it). Plausibly, such chanciness rules out knowledge-wh. I think we can multiply cases like T-CELL with similar results, which in turn supports the following general principle about knowledge-wh for two-alternative questions:

Equally Likely Alternatives (ELA)
>Necessarily, UT-processes do not generate knowledge-wh.

Next, note how questions with numerous alternatives have a corresponding two-alternative question that occupies the very same logical space. For example, suppose there are only 91 people who have access to the locker room—Steve the groundskeeper and the 90

members of the football team. Here, we can meaningfully ask two distinct questions that correspond to the same logical space:

> "Who left the locker room door unlocked—Steve, or Jim, or Bill, or Joe, or…etc.?"

> "Who left the locker room door unlocked—Steve, or someone from the football team?"

With respect to the latter question, the logical space for the alternative, *someone from the football team,* is identical to the disjunction of logical spaces corresponding to *Jim*, *Bill*, *Joe*, and all the other members of the team. We can easily multiply cases with pairs of questions like these. It seems that for *any* multi-alternative question, there's a corresponding two-alternative question that we can meaningfully ask as well. The following principle captures this thought:

> Two-alternative Analog (TA)
> Necessarily, for any question Qa that has p as an alternative, there exists some question Qb with p as an alternative that both denotes the same logical space as Qa *and* has only one other alternative to p.

Now, let's examine a further variant on COLD in which it's clear that there are two distinct questions Tim could answer, where one of the questions is the two-alternative analog of the other.

> COLD/T-CELL
> All of the details are the same as COLD regarding the statistical probabilities of H_1-H_{101} across cases of D_S, except here, N also specifies that whenever H_1 causes D_S symptoms in a patient, the D_S symptoms always come along with an elevated T-cell count. In contrast, N states that whenever any of H_2-H_{101} cause D_S, that instance of D_S never comes along with an increased T-cell count in the patient. Tim knows all of this to be the case as he observes patients with D_S.

We can distinguish the following two questions: "does the patient with D_S symptoms have an elevated or normal T-cell count?" and "which condition causes the patient's D_s symptoms— H_1 (the cold) or H_2 or…H_{101}?" Also, in COLD/T-CELL, the logical space corresponding to

the H_1 alternative is equivalent to the logical space corresponding to the elevated T-cell count alternative, and Tim is aware of this fact.

Here, let's consider this question: if Tim knows *the patient's D_S symptoms are caused by H_1 rather than any of H_2-H_{101}*, is Tim also in a position to know *the patient's T-cell count is elevated rather than normal?* Intuitively, it seems that he must be in such a position. Imagine Tim stating the following conjunction: "I know the patient's symptoms are caused by the cold and I know that their being caused by the cold comes with an increased T-cell count, but I don't know whether this patient has an increased or normal T-cell count." Intuitively, something's gone wrong; this statement is akin to any of the more absurd Moorean conjunctions. This case, and the fact that we can easily multiply cases like it, provides support for a general closure principle for knowledge-wh:

> Process Question Closure (PQC)
> > Necessarily, for any two questions Qa and Qb that both include the alternative p and where Qb only has two alternatives [p,q]
> > > If,
> > > > -S uses method P and thereby comes to know (i) *p is the answer to Qa*
> > > > -S knows (ii) *the alternatives of Qa and Qb occupy the exact same logical space*
> > > Then,
> > > > -S is in a position to know *p is the answer to Qb* (by competently deducing this from (i) and (ii))

In his own discussion of knowledge-wh expansion by entailment, Schaffer nicely articulates an underlying reason behind a principle like PQC. He considers two questions, "Q" and "q" that occupy the very same logical space and that both have p as one of their alternatives (2007b: 251, fn. 17). He stipulates that "Q partitions the contrasts" to p into multiple alternatives but that q "lumps" the contrasts to p "into one big disjunction" (251). Schaffer acknowledges that the state of *knowing p is the answer to Q* "contains more information" than the state of *knowing p is the answer to q* given that Q includes more partitions the relevant logical space (251). But Schaffer thinks that this additional information in Q, by itself, "*plays*

11

no epistemic role. All the alternatives must be eliminated, *however they are partitioned*' (251., emphasis mine). This seems exactly right, as cases like COLD/T-CELL illustrate. In order to know p is the answer to some question, the process one uses must enable the subject to eliminate *the entirety* of the logical space occupied by p's competitors. Whether or not the entirety of this logical space is partitioned into smaller sub-spaces, by itself, seems to make no epistemic difference to whether some process enables the subject to eliminate the entirety of this space.

As I'll argue, PQC, in combination with ELA and TA, rules out the possibility that subjects—like Tim in COLD—could come to know an answer to a question by using a UBA process.

Let's assume, for reductio, that a process like P_T does allow Tim to know H_1 is the answer to Q_C in COLD. Now, given TA, there's a corresponding two-alternative question Q_C' that Tim could answer instead. Q_C' corresponds to the same logical space as Q_C, includes H_1 as one of its alternatives, and has the disjunction of H_2-H_{101} as its second alternative. According to ELA, Tim can't use method P_T to come to know *H_1 is the answer to Q_C'*, given that Q_C' is a two-alternative question and, relative to P_T, H_1 is just as likely as Q_C''s sole other alternative. But if our assumption (for reductio) is correct, then Tim *can* use P_T and thereby come to know *H_1 is the answer to Q_C*. And, given PQC, Tim could then competently deduce, and thereby come to know, *H_1 is the answer to Q_C'*.

Yet, this is a puzzling and seemingly absurd result. There doesn't appear to be an epistemologically relevant difference between *directly* using method P_T to answer Q_C' on the one hand and deducing an answer to Q_C' from one's prior usage of P_T (for answering Q_C) on

the other. After all, *both* of these processes for answering Q_C' count as UT-processes.[5] If knowledge gained from directly using method P_T to answer Q_C' is a violation of ELA, then knowledge gained from its indirect, deductive usage surely violates ELA as well.

Considering COLD/T-CELL further illustrates this point. In COLD/T-CELL, it's implausible that Tim comes to know *the patient has an increased T-Cell count rather than a normal T-Cell count* were he to (correctly) deduce this conclusion from his answer *the patient's D_S symptoms are caused by the cold, not by H_2, H_3, …nor H_{101}* arrived at through the use of P_T. After all, this deductive procedure is a UT-process given the 50/50 statistical ratio of increased T-cell counts to normal T-cell counts in D_S patients. But if this is right, then it's doubtful that using P_T in COLD/T-CELL allows Tim to know *the patient's D_S symptoms are caused by the cold, not by H_2, H_3…, nor H_{101}* in the first place. Moreover, there doesn't appear to be any epistemologically relevant difference between using method P_T to answer Q_C in the original COLD case and using P_T to answer the question, *what causes the patient's D_S symptoms—H_1, or H_2, or…H_{101}?* in COLD/T-CELL.

Recall, the dubious result that we're considering is as follows: in COLD, Tim *can* come to know H_1 is the answer to Q_C', by deducing this from the answer to Q_C acquired from P_T, despite being unable to know H_1 is the answer to Q_C' through the direct use of method P_T. To avoid this conclusion, one could simply reject PQC. But PQC is very plausible; giving up PQC seems highly problematic. Next, one might simply deny ELA, which then opens the possibility that directly (or indirectly) using P_T *can* allow Tim to know the answer to Q_C' in COLD. But as we saw, ELA is also very reasonable. In essence, ELA captures our clear

---

[5] Relative to the more indirect procedure—that first uses P_T to answer Q_C and then delivers an answer to Q_C' through deduction—it's still the case that H_1 is just as likely as the sole other competitor for Q_C'. After all, in answering both Q_C and Q_C' with H_1, the same logical space gets eliminated, and the indirect-deductive procedure that yields the answer to Q_C' simply uses the prior answer to Q_C as a premise.

intuition that when two alternatives are equiprobable (relative to the subject's method), selecting the *correct* alternative would be a matter of chance—a kind of chanciness that's inconsistent with gaining knowledge. At the very least, one who denies ELA takes on the burden of *either* giving an error theory for our intuitions of knowledge-undermining chanciness in cases like T-CELL, *or* explaining why such chanciness undermines knowledge in T-CELL but somehow doesn't in cases like COLD. At this point, it's unclear where such explanations would even begin. I contend that the most reasonable option is to simply relinquish our initial (reductio) assumption—namely, that using P_T can allow Tim to know H_1 is the answer to the original question Q_C in COLD.

According to TA, there's a two-alternative analog to *any* question with numerous alternatives. Hence, for any process with respect to which the target answer and the disjunction of alternatives are equiprobable, we can conceive of a corresponding (deductive) UT-process for answering the two-alternative analog of that question—which in turn generates the very same kind of puzzle associated with answering Q_C and Q_C' considered above. This suggests that ELA, TA, and PQC are jointly inconsistent with the idea that knowledge-wh can be generated by processes for which the target answer and the disjunction of alternatives are equiprobable. In essence, the first part of my closure argument for a reliability condition on knowledge-wh is simply a statement of this inconsistency.

W1. PQC, ELA, and TA.
W2. If it's possible for a subject to gain knowledge-wh from a process P, relative to which the P's selected answer is equally or less likely than the disjunction of the other alternatives, then either ELA, TA, or PQC is false.
W3. Therefore, it's impossible for a subject to gain knowledge-wh from a process P, relative to which P's selected answer is equally or less likely than the disjunction of the other alternatives.

14

While W3 establishes that UBA processes—like P_T—cannot generate knowledge-wh, it does not, strictly speaking, state a reliability condition on knowledge-wh. But it does entail such a condition once combined with the following premise.

> W4. Necessarily, if a subject's method P for answering a question is unreliable, then with respect to P, P's selected answer is equally or less likely than the disjunction of the question's other alternatives.

In defense of W4, let's consider an unreliable process P_x that selects answer p_x to question Q_x. Since P_x is unreliable, at least 50% of the time p_x doesn't characterize reality, and some alternative state of affairs obtains instead. Plausibly, these alternative ways that reality could turn out—in the epistemically relevant sense—*just are* the other Q_x competitors to p_x. But if this is right, then P_x's being unreliable entails that the disjunction of p_x's competitors is at least as likely as p_x itself. This case is clearly schematic, which gives us good reason to accept the general principle W4. W3 and W4 entail reliabilism about knowledge-wh, which we can capture as follows:

> W5. (From W3 and W4) Necessarily, if a subject uses an unreliable method P to answer a question, the subject does not gain knowledge-wh.

Crucially, W1-W5 in no way presupposes that a process's salient knowledge-determining feature is its *likelihood of generating truth over falsehood.* Instead, this argument only invokes a reasonable knowledge-wh closure principle and the idea that a process generates knowledge-wh only if the target answer is more likely than each of its alternatives. Together, I take R1-R3 and W1-W5 to provide grounds for reliabilism about knowledge *in general.* In the next two sections, I'll show that neither of Turri's arguments for unreliable knowledge succeed in undermining the premises of the two reliabilist arguments presented above.

3. The Argument from Explanatory Inference

3.1 House and Unreliable Knowledge

Turri thinks that at least some cases of inference to the best explanation provide us with counterexamples to reliabilism:

> Inference to the best explanation yields knowledge if the explanation we arrive at is true. But even when it is true, the best explanation might not be very likely. So, our disposition to infer to the best explanation might not be reliable. So unreliable knowledge is possible. (2015: 536)

Turri offers two distinct reasons to think that unreliable inferences to the best explanation can produce knowledge. First, as I discussed in §2.2, Turri notes how some inferences to the best explanation fit the general form of a UBA process. In his view, it just seems that UBA processes can generate knowledge (537).

In addition to these general considerations, Turri also presents what he takes to be a specific "case study" of knowledge from unreliable inference to the best explanation: the character House from the hit TV show named after him (537). In the series, House is a specialist in rare-disease diagnosis at a teaching hospital. Turri describes the plot as follows:

> Most episodes unfold similarly. The patient presents with symptoms that House finds "interesting" enough to investigate. House's team then deliberates, makes a diagnosis, prescribes a course of treatment which fails, revisits the matter in light of the failed treatment, new information, or a change in symptoms, then issues another diagnosis, prescribes a new course of treatment which fails, revisits the matter in light of the failed treatment, new information, or a change in symptoms, etc. This cycle continues until they finally solve the case and save the patient's life…House and his team explicitly reason abductively…For present purposes, a crucial aspect of the series is that, in the end, House knows what disease the patient has. And he knows despite being unreliable. (537-8)

Turri formulates the above reasoning into the following argument:

E1.   If House knows, then unreliable knowledge is possible. (Premise)
E2.   House knows. (Premise)
E3.   So unreliable knowledge is possible. (From 1 and 2) (539)

In support of E1, Turri states that House's belief-forming method has the critical feature of a UBA process in the following sense: as demonstrated by the substantial track record of

prior diagnostic failure leading up to the correct diagnosis, House's diagnostic method has a likelihood of generating falsehood that's 50% or higher (537). Turri also points to an assortment of direct quotations from the show indicating that the other characters think House's methods are unreliable (537-8). Despite House's diagnostic unreliability, Turri contends that E2 is our intuitive reaction at the end of each episode when House finally lands on the correct diagnosis (538). If E1 and E2 are correct, then we can deduce that unreliable knowledge is possible.

3.2   A reply to the argument from explanatory inference

In considering the general form of UBA processes, we can acknowledge that there's some plausibility to the idea that these processes generate knowledge. However, the reasons in favor of W1-W3 would seem override whatever initial intuitive pull such considerations might have had in favor of unreliable knowledge.[6] Hence, I take it that the argument from explanatory inference hangs on Turri's specific counterexample to reliabilism.

While House's diagnostic procedure is an interesting case, I'll argue that it ultimately fails to establish the possibility of unreliable knowledge and that E1 remains unmotivated. Even granting that House's correct diagnoses count as knowledge, there's no reason to think that such diagnoses stem from unreliable processes. Turri's defense of E1 founders on how we ought to categorize, or "type" House's belief-forming process. To see this, let's turn our attention to the famed generality problem for reliabilist epistemology. The generality problem is simply the challenge of providing an account of a belief-forming process *token's* epistemically relevant process *type*. Importantly, belief-forming process tokens instantiate innumerable

---

[6] See fn. 18 below for a potential explanation for why we're (mistakenly) drawn to the idea that UBA processes can generate knowledge.

types. Moreover, a token's various process types will correspond to different degrees of reliability.[7] So, we need some grasp of the epistemically relevant process type whose reliability score determines whether the token process in question generates knowledge.[8]

For this reason, Turri's defense of E1 hinges on *which* type is epistemically relevant for House's token processes that end up delivering true diagnoses. In reconstructing Turri's reasoning, he seems to entertain *two* distinct relevant-type candidates for House's token processes, which in turn produces to two distinct defenses of E1. On the one hand, Turri suggests that the relevant type is something broad like *inference to the best explanation.* On the other hand, Turri seems to indicate that the relevant type is much narrower, such that House's diagnostic method counts as a UBA process involving a specific set of empirical data. But as I'll argue below, neither approach to typing House's process delivers a viable defense of E1.

First off, let's consider Turri's claim that House's "relevant method" is "inference to the best explanation" (539). With this relevant type assignment, Turri defends E1 by pointing out that, in using abduction, "[u]sually [House] gets it wrong at least two or three times before finally getting it right" (538). While we can grant that House's track record with the process type *inference to the best explanation* is mediocre at best, we have good independent reason to doubt that House's relevant process type is something so broad as *inference to the best explanation.*

As it turns out, a very broad, coarse-grained approach to typing belief-forming processes runs counter to several of the more promising responses to the generality problem. Importantly, each of these promising responses identifies relevant types as being much

---

[7] For example, a given token instance of seeing (and believing) that a car is coming will instantiate numerous process types, including *visual cognition* and *visual cognition under good lighting conditions.* Clearly, the latter process type is more reliable than the former. Richard Feldman (1985) and Goldman (1979) are key figures who highlighted the importance of the type/token distinction for making sense of reliabilism—and for setting up the generality problem.
[8] See Conee and Feldman (1998) for a classic statement of the generality problem and a critical survey of numerous attempts to answer the generality problem. Notably, Turri does briefly address the generality problem as he responds to a concern raised by Heather Battaly (2015: 539).

narrower—building in detailed descriptions of both the *content* of the target belief produced by the token process and the *evidence* (or, grounds) on which that belief is based in the token process.[9] For short, let's call this general strategy the *narrow content-evidence response* to the generality problem. So, according to this response, the relevant type for Tim's process in COLD is the following narrow content-evidence pair:

P_T [Tim's judging whether the patient has H_1, on the basis of data set N and observing the patient's D_S symptoms]

Let's briefly survey some of the reasons in favor of the narrow content-evidence response. At first glance, the target belief's content and the evidence on which it's based naturally suggest themselves as being epistemically relevant—in contrast to other features of a token process, like *what the subject was wearing when she formed the belief,* or *whether the resultant belief makes one happy or sad,* etc. Secondly, by typing processes narrowly and factoring in the total evidence used by the subject, this solution to the generality problem accommodates the intuitive idea that even tiny changes in one's body of evidence can have a tremendous impact on the epistemic status of the target belief. For instance, in cases of reasoning abductively with complex bodies of data, sometimes acquiring one small piece of evidence e_x can take a formerly unclear question and "tie everything together" in a way that identifies a clear best answer. In such a case, it's intuitive that an abductive inference that incorporates evidence e_x is much more reliable than an abductive inference that doesn't. But if epistemically relevant types were as broad as *inference to the best explanation,* this wouldn't be the case, as both abductive inference tokens would fall

---

[9] Early in the literature on reliabilism and the generality problem, Alvin Goldman defended the thesis that the relevant type is "the *narrowest* type that is *causally operative* in producing the belief token in question" (1986: 50). The idea that relevant types are narrow content-evidence pairs has recently been defended by Juan Comesaña (2006:37) and me (2021: 5634-5640; 5642-3), both drawing on earlier arguments made by William Alston (1995: 27).

under the very same process type. Hence, it seems that relevant types must factor in detailed descriptions of the evidence used throughout the token process.

If the narrow content-evidence response to the generality problem is on track, then assigning such a broad type, like *inference to the best explanation,* to House's diagnoses is mistaken. As a result, the defense of E1 that relies on typing House's diagnostic processes so broadly is unsuccessful.

As I noted above, Turri also suggests that House's procedure, in which he infers the most likely diagnosis given the *particular* empirical evidence available, is a UBA process. Since this way of categorizing House's process builds in specific details about the evidence in use, the relevant type will be much narrower than *inference to the best explanation* and more similar to types like P_T. On the one hand, this way of typing House's procedure comports much better with an independently plausible solution to the generality problem. However, for House's token processes that finally arrive at the correct diagnosis*,* we have little to no reason for thinking that their corresponding narrow types count as UBA processes.

Remember, at the beginning of each episode, House identifies a short list of diagnoses that are the most promising (i.e., most likely) explanations of the patient's symptoms. House and his team then start testing these diagnoses one by one. This being the case, there are two crucial ways in which House's evidential situation changes throughout the course of the episode. First, as House and his team rule out competing alternative diagnoses throughout an episode, the correct diagnosis naturally comes to occupy a greater percentage of the relevant probability space. Second, in some episodes, House and his team acquire additional pieces of empirical evidence that go well beyond the symptoms presented by the patient at the beginning of the episode. For example, more than halfway through the episode "Occam's Razor," House learns a new piece of information about the *sequence* in which the patient's various symptoms

arose.[10]   This ends up being a watershed piece of evidence that substantially increases the probability of the correct diagnosis.

Given these two sorts of changes in House's evidential situation throughout the course of an episode, it's doubtful that the narrow types for House's knowledge-generating belief-forming process tokens (at the end of each episode) are UBA processes.   While the correct diagnosis might have had a probability of .5 or less given the evidence House *starts with*, there's no reason to believe that this likelihood remains at .5 conditional on the evidence House possesses towards the end of an episode.   According to the narrow content-evidence response to the generality problem, it's House's *updated* evidential situation that's built into the relevant type for his knowledge-generating process token.   Thus, insofar as Turri's defense of E1 relies on the idea that the relevant types for House's knowledge-generating diagnostic tokens are UBA processes, this defense is unsuccessful.

As we've seen, there are problems with both of Turri's arguments for E1.   House might be unreliable with respect to *inference to the best explanation,* but it's doubtful that this broad type is epistemically relevant.   When we type House's process tokens more narrowly, there's no reason for thinking that he arrives at correct diagnoses unreliably.   As a result, E1 remains unsupported.

4. The Argument from Achievements

4.1   Achievement and Unreliable Ability

The argument from achievements turns on the idea that knowledge is a kind of achievement—namely, an intellectual achievement.   According to Turri, characterizing knowledge as an

---

[10] Season 1-episode 3. This revelation occurs 38 minutes into the episode.

achievement ought to make one doubtful of reliabilism. He formulates the argument from achievements as follows:

A1.   Achievements don't require reliable abilities. (Premise)
A2.   If achievements don't require reliable abilities, then unreliable knowledge is possible. (Premise)
A3.   So unreliable knowledge is possible. (From A1 and A2) (2015: 531)

In defense of A1, Turri presents a host of straightforward achievements that arise from unreliable abilities. When a professional baseball player gets a hit, this achievement stems from ability even though the most skilled major league hitters manage to get hits in only one third of their at-bats. The same can be said about professional hockey goalscoring—where even the best players only manage to score goals once out of every eight shots (531). Lastly, Turri states that novices at a given task can still successfully achieve the goal of that task despite being unreliable at that task (given their current status as a novice) (531).

Importantly, Turri clarifies that A1 should be read as stating a dominant tendency of the category *achievements*. This is similar to how the phrase, "humans don't have eleven fingers" is used (correctly) to describe the way humans *typically* are as opposed to stating a universal generalization for all humans (534). In addition, Turri specifies that A2 doesn't state a material or strict conditional but should instead be interpreted in the following way: "[k]nowledge is an intellectual achievement, so absent a *special reason* to think otherwise, we should expect it to share the profile of achievements generally" (532, emphasis mine). Importantly, Turri goes on to defend the idea that we lack such a "special reason to think otherwise" by rebutting a handful of arguments purporting to show that the knowledge achievement can only arise from

reliable abilities.[11] Insofar as these arguments fail, Turri concludes that A1-A3 provides a compelling case for unreliable knowledge.

4.2   A reply to the Argument from Achievements

My response to this argument is two-fold. First, I present an interpretive challenge to A1. As it turns out, it's not at all clear how we're to understand the phrase "require reliable abilities." Given this unclarity, I'll demonstrate the difficulty of identifying a reading of A1 that's *both* plausible *and* succeeds in picking out a dominant tendency of achievements that's violated by reliabilist knowledge.   Perhaps this challenge can be overcome; perhaps not. Regardless, I also argue that we *do* possess "special reasons" for thinking that knowledge is distinct from other achievements in having the specific sort of reliability condition that reliabilists posit—a result that undercuts A2. Hence, the idea that reliabilist knowledge violates a dominant tendency of achievements in a way that's *ultima facie* problematic (for reliabilism) remains unmotivated.

Let's start with A1.   Presumably, in this context we're understanding an ability to be, at least partially, individuated by the kind of *outcome* that the ability aims to bring about. This way of thinking about ability allows us to sketch our first candidate interpretation of an achievement's "requiring reliable abilities:"

> RR₁   Achievement A requires reliable ability just if, in order to achieve A, the subject S must possess an ability that reliably brings about A.

---

[11] Here, I'll briefly canvass some of these arguments and Turri's responses. First, one might think that knowledge is *valuable,* and that in order for an achievement to be valuable, it must stem from a reliable ability. But Turri correctly points out that other genuinely valuable achievements—like a major league hit—come from unreliable abilities (533). Next, perhaps knowledge is a *creditable* achievement*,* and agents can receive credit for their achievement only if they used a reliable ability. Once again, however, the professional hockey player *is creditable* for scoring a goal even though he's very unreliable when it comes to scoring goals during games (533).

RR$_1$ nicely captures Turri's examples of "unreliable achievement," which in turn supports A1's assertion that achievements in general have the dominant tendency of *not* requiring reliable abilities in the sense of RR$_1$. However, it's doubtful that reliabilist knowledge violates this dominant tendency. As an anonymous referee points out, reliabilism does not, strictly speaking, claim that one must use an ability or process that *reliably delivers knowledge* in order to gain knowledge. Instead, the reliability condition only states that, in order to gain knowledge, one must use an ability that reliably delivers *true belief*.

This point suggests a second reading of "require reliable abilities," on which reliabilist knowledge plausibly *would* violate the dominant tendency expressed in A1. While the outcome of a true belief doesn't suffice for knowledge, true belief *is* a non-trivial and essential constituent of knowledge. Perhaps we're to understand an achievement A's requiring a reliable ability as follows:

> RR$_2$ Achievement A requires reliable ability just if, in order to achieve A, the subject must possess an ability that reliably brings about at least one non-trivial constituent of A.[12]

However, insofar as we interpret A1 as stating a dominant tendency for achievements on the RR$_2$ interpretation, then A1 looks doubtful. Upon consideration, it seems that many achievements *do* require the possession of reliable abilities to bring about non-trivial constituents of those achievements. For example, successfully achieving a baseball hit is partially constituted by successfully gripping a bat, swinging a bat, visually tracking a moving object, hitting a moving object with a bat, just to name a few. Moreover, it seems that a subject can achieve a hit only if she can *reliably* bring about at least some of these non-trivial

---

[12] One might think that achievements like *breathing* or other tasks related to staying alive are constituents of achievements like *hitting a baseball*. Even if this is right, there's a clear sense in which *breathing* is trivially constitutive of *hitting a baseball*—since it's constitutive of virtually anything a human does—in a way that *swinging a bat* is not.

constitutent outcomes. When a major league baseball player has a bad season and only gets hits in 5% of his at-bats, it's still a genuine achievement on those (rare) occasions when he does get a hit. But importantly, it seems that these count as achievements only because his hits manifest a host of abilities that are reliable for him. For instance, whenever he tries to swing the bat at the ball, he can—with almost 100% reliability—execute the swing. Also, consider his noteworthy reliable ability to *hit a ball with a bat*. Despite being somewhat poor at hitting 95 mile-per-hour fastballs during a game, when he's swinging at 77 mile-per-hour pitches in batting practice, he'll crush almost every single ball at least 250 feet through the air—a feat that's virtually unthinkable for most humans.

Contrast this case with a subject who's *much* less coordinated and focused. When this fellow tries to hit a pitch, roughly 50% of the time his efforts go horribly awry; either he drops the bat on the ground, or he loses sight of the ball completely, or swings at some object other than the ball. In the other 50% of his attempts, he actually executes a swing. Lo and behold, in 5% of his total attempts, he gets a hit. Now, while it's clear that the hits of the slumping major leaguer are genuine achievements, it's by no means clear that the hits of this woefully uncoordinated fellow constitute achievements. While both men hit the ball at the same statistical frequency, the hits of the uncoordinated man are marked by a significant degree of chanciness that doesn't characterize the major leaguer's hits. At the very least, even *if* the uncoordinated man's hits are achievements of some sort, it seems incorrect to classify them under the same *kind* as the major leaguer's achievements. On reflection, this is because the major leaguer possesses an important sort of *control* over the outcome of his attempts that the uncoordinated man lacks. This control seems at least partially constituted by the major leaguer's reliability with respect to the constitutive abilities discussed above. I think we can

easily multiply examples like this for other types of achievements.[13] As a result, we have good reason to think that the RR_2 sense of "require reliable abilities" *is* a dominant tendency of achievements.   Hence, on the RR_2 reading, premise A1 is unmotivated.[14]

The argument from achievements needs a different interpretation of "require reliable abilities," and comments from Turri shed light on how this might go. When he originally describes the hitting achievements of baseball star Ted Williams, Turri states, "The *relevant* ability could at best be counted on to produce a hit about four in ten times" (531, emphasis mine). This statement suggests that a given kind of achievement A has a corresponding *relevant* ability that's determinative of whether an event counts as a genuine instance of A. With this understanding of achievements in mind, we can frame a corresponding interpretation of "require reliable abilities" as follows:

> RR_3 Achievement A requires a reliable ability just if, in order to achieve A, the subject S must be reliable with respect to A's relevant ability.

On the RR_3 reading of the argument, premise A1 states that achievements have the following dominant tendency: subjects needn't be reliable with respect to achievement A's relevant ability in order to genuinely achieve A. This gives us a reason to reject reliabilism if we suppose that the *ability to obtain true belief* is the relevant ability for the knowledge achievement.

But can we even make sense of this notion of *ability relevance* that applies across all sorts of different achievements? Whatever it amounts to, an achievement's "relevant ability" can't just mean something like "the single ability, corresponding to a given achievement A, whose success rate determines whether an event is a genuine instance of A or just an event of sheer chance." As the case of the uncoordinated man illustrates, for many achievements, the success

---

[13] For instance, achieving a *made shot* in a basketball game arguably requires a reliable ability to *lift the basketball, aim the basketball, toss the basketball with the correct trajectory,* etc.

[14] Much thanks to the anonymous referee who made several helpful suggestions and encouraged me to develop this line of response.

rates of *several* abilities—not just one—play essential roles in determining whether that achievement occurs.[15] But then what is it that makes a particular ability *the* relevant one (in Turri's sense) for a given achievement? Can we identify any sort of role this concept of relevance might play? Moreover, do we have any reason to think such a concept applies to all different kinds of achievements?[16]

I won't argue that defenders of unreliable knowledge *couldn't* develop answers to these questions. But these difficulties cast doubt on whether achievements (in general) even have the kind of dominant tendency that's problematic for reliabilism—problematic in the sense that we lack special reasons for thinking knowledge could violate that tendency. These doubts become more acute as one considers the ways in which we possess special reasons to think that knowledge deviates from other dominant tendencies of achievements.

On this point, it's instructive to examine a kind of parity argument to A1-A3. Recently, there's been much debate on whether one's prior knowledge can be defeated by "higher-order considerations" like the following: good evidence that one's belief is false, good evidence that one's belief lacks justification, good evidence that one's belief-forming process doesn't meet whatever reliability threshold knowledge requires, etc. Otherwise put, this is a debate over whether there's a no-higher-order defeater condition on knowledge.[17] Ultimately, I find this debate interesting and substantive, with powerful considerations on either side of the issue.

---

[15] Reliabilists are only committed to the thesis that a token process generates knowledge only if its *epistemically* relevant ability/process type is reliable, which many reliabilists take to be the token's narrow content-evidence pair. This thesis is compatible with the idea that success-rates for *other* process types belonging to the token process play some determinative role in whether the token produces knowledge.

[16] I think we can reasonably identify narrow content-evidence pairs as being the *epistemically* relevant ability/process type that figures in the reliability condition on knowledge. However, it's not at all clear that this identification stems from our grasp of a relevance concept that applies to other kinds of achievements besides knowledge.

[17] See Goldberg and Matheson (2020) for a defense of a no-higher-order defeater condition on knowledge. Also, see Lasonen-Aarnio (2010) for a prominent objection to such a condition.

Notably, we can identify a dominant tendency of achievements that's pertinent to this debate. It seems that most achievements fit the following pattern:

> EA   Subject S can achieve A even though S possesses strong evidence that *S doesn't achieve A.*

For example, we can easily imagine a very talented quiltmaker who has, unfortunately, acquired remarkably strong (yet totally misleading) evidence that he both lacks the abilities necessary for being a good quiltmaker and that lucky artistic outcomes are pervasive throughout his lifetime. So, whenever he completes his amazing quilts, he believes that he hasn't achieved anything of value—and his evidence supports this belief. But despite his own viewpoint, his quilts clearly are his achievements. We can multiply cases like this for all sorts of other achievements, which supports the claim that EA is a dominant tendency of achievements.

Now notice, the higher-order defeat considerations mentioned above *just are* different ways of having evidence that one hasn't achieved knowledge. So, let's consider—does the fact that EA is a dominant tendency for achievements provide us with a compelling reason to reject a no-higher-order defeater condition on knowledge? On reflection, I think the answer is clearly *no.* The interesting arguments in favor of a no-higher-order defeater condition don't seem to be significantly undermined or overridden by the fact that people can genuinely achieve things like baseball hits and quilt-making despite having evidence that they don't actually achieve such things. This suggests that we understand knowledge to be *sufficiently different* from other sorts of achievements in a way that undercuts any inference to the conclusion that knowledge satisfies EA simply because many other achievements satisfy EA. Without getting too technical, I think we can characterize this understanding as follows: we grasp that part of the essence of knowledge—and what gives knowledge its unique value—is a matter of how the subject's *evidence* furnishes her with a viewpoint on the world or herself. Clearly, this is nothing like the essence of quilt-making or baseball hitting.

Now, let's assume for the sake of argument that we can somehow specify the relevant sense of "requires reliable abilities" such that reliabilist knowledge violates a dominant tendency of achievements. Much like what we saw with knowledge and the EA dominant tendency, I contend that our understanding of knowledge also provides us with special reason to think that knowledge is distinct amongst the other achievements in having the specific kind of reliability condition that reliabilists defend. Interestingly enough, the reliabilist arguments from §2 nicely capture these reasons.

First, it's highly plausible that binary knowledge is closed under entailment relation PC and that knowledge-wh is closed under entailment relation PQC. Secondly, there seems to be a clear sort of knowledge-precluding chanciness that comes with carrying out the PC deduction when the antecedent is adopted through an unreliable process. A similar sort of chanciness characterizes any UT-process, which in turn shows that we can only affirm the possibility of unreliable knowledge-wh on pain of denying PQC. Notice, there's no sense in which baseball hits and quilt production are "closed" under any sort of entailment relation. Moreover, given the nature of *epistemically relevant* process/ability types (as discussed in §3), whether or not a process delivers true belief with knowledge-precluding chanciness is a function of how the subject's *specific evidence* determines the likelihood of the target belief's *content*. There's no analogous function involved in tasks like baseball hitting. This being the case, there doesn't seem to be any reason for thinking that the norms governing knowledge-precluding chanciness and the norms governing hit-precluding chanciness would fix the *same* minimal success rates required for avoiding each respective sort of chanciness.

To sum up, R1-R3 and W1-W5 provide us with independent grounds to think that knowledge deviates from whatever dominant tendency of achievements that might otherwise count against reliabilism. This result, coupled with the fact that we've yet to identify the

relevant sense in which achievements don't "require reliable abilities" to begin with, indicates that the conjunction of A1 and A2 remains unmotivated.[18]

Conclusion

I began by addressing a lacuna in the reliabilist program by formulating a new independent argument for reliabilism. As it turns out, simple facts about knowledge-closure support reliabilism about knowledge-that and knowledge-wh in way that goes well beyond a simple appeal to intuitive plausibility. Furthermore, despite the ingenuity of the argument from explanatory inference and the argument form achievements, I've shown that neither argument undermines the closure-based argument for reliabilism. As things stand, the reliabilist orthodoxy remains intact.[19]

References

Alston, William. (1995), "How to Think about Reliability." *Philosophical Topics*. 23.1: 1-29.

Comesaña, Juan. (2006), "A Well-Founded Solution to the Generality Problem." *Philosophical Studies*. 129.1: 27-47.

Conee, Earl and Richard Feldman. (1998), "The Generality Problem for Reliabilism." *Philosophical Studies*. 89.1:1-29.

---

[18] Far from undermining reliabilism, the fact that knowledge is an achievement serves to undercut some of the intuitive plausibility of unreliable knowledge. As I noted in §2.2 and §3.2, there is some intuitive pull towards ascribing knowledge in cases where subjects successfully use UBA processes. But if knowledge is an achievement, we can sketch an error theory for how these intuitions go astray. A reliabilist can grant that true beliefs delivered by UBA processes count as genuine intellectual achievements, because reliabilism doesn't claim that knowledge is the only intellectual achievement involving true belief. On this point, see Pritchard (2012: 248, fn. 5; 258) where he defends the view that some true beliefs that fall short of knowledge are genuine intellectual achievements (i.e., successes through ability). Now, we can sketch the reliabilist-friendly error theory as follows: while we're (correctly) sensitive to the presence of epistemically valuable intellectual achievement in truth-yielding UBA processes, the fact that knowledge is the most prominent (and often-ascribed) intellectual achievement in everyday discourse contributes to why we're mistakenly inclined to categorize these achievements as *knowledge* achievements. Much thanks to an anonymous referee who encouraged me to address this issue.

[19] Much thanks to Ernie Sosa, Matt McGrath, Chris Willard-Kyle, Andrew Moon, Carolina Flores, and Chris Copan for several helpful comments on earlier drafts of this paper.

Feldman, Richard. (1985), "Reliability and Justification." *The Monist*. 68.2: 159-174.

Goldberg, Sanford and Jonathan Matheson. (2020), "The Impossibility of Mere Animal Knowledge for Reflective Subjects. *Erkenntnis*. 85.4: 829-40.

Goldman, Alvin. (1979), "What is Justified Belief?" *Justification and Knowledge*. Ed. by G.S. Pappas. Dordrecht: D. Reidel Publishing. 1-23.

——— (1986), *Epistemology and Cognition*. Harvard: Harvard University Press.

Lasonen-Aarnio, Maria. (2010), "Unreasonable Knowledge." *Philosophical Perspectives*. 24.1: 1-21.

Pritchard, Duncan. (2009), "Apt Performance and Epistemic Value." *Philosophical Studies*. 143:407-416.

——— (2012), "Anti-luck Virtue Epistemology." *The Journal of Philosophy*. 109.3: 247-279.

Roeber, Blake. (2020), "Is Every Theory of Knowledge False?" *Noûs*. 54.4: 839-866.

Schaffer, Jonathan. (2004), "From Contextualism to Contrastivism." *Philosophical Studies*. 119: 73-103.

——— (2007a), "Knowing the Answer." *Philosophy and Phenomenological Research*. 85: 383–403.

——— (2007b), "Closure, Contrast, and Answer." *Philosophical Studies*. 133: 233-55.

Sosa, Ernest. (2007), *A Virtue Epistemology: Apt Belief and Reflective Knowledge*. Oxford: OUP.

Tolly, Jeffrey. (2021), "Knowledge, Evidence, and Multiple Process Types." *Synthese*. 198: 5625-5652.

Turri, John. (2015), "Unreliable Knowledge." *Philosophy and Phenomenological Research*. 90.3:529-545.

——— (2016), "A New Paradigm for Epistemology: from Reliabilism to Abilism." *Ergo: an Open Access Journal of Philosophy*. 3.8: 189-231.