ORIGINAL RESEARCH

Anti-luminosity and anti-realism in metaethics

Jussi Suikkanen[1] 🔘

Abstract

This paper begins by applying a version of Timothy Williamson's anti-luminosity argument to normative properties. This argument suggests that there must be at least some unknowable normative facts in normative Sorites sequences, or otherwise we get a contradiction given certain plausible assumptions concerning safety requirements on knowledge and our doxastic dispositions. This paper then focuses on the question of how the defenders of different forms of metaethical anti-realism (namely, error theorists, subjectivists, relativists, contextualists, expressivists, response dependence theorists, and constructivists) could respond to the explanatory challenge created by the previous argument. It argues that, with two exceptions, the metaethical anti-realists need not challenge the argument itself, but rather they can find ways to explain how the unknowable normative facts can obtain. These explanations are based on the idea that our own attitudes on which the normative facts are grounded need not be transparent to us either. Reaching this conclusion also illuminates how metaethical anti-realists can make sense of instances of normative vagueness more generally.

✉ Jussi Suikkanen
    j.v.suikkanen@bham.ac.uk

1   Department of Philosophy, University of Birmingham, B15 2TT Birmingham, UK

🖄 Springer

1 Introduction

A lively metaethical debate about normative vagueness has been conducted recently mainly from the perspective of metaethical realism.[1] Metaethical realists have used normative vagueness to explain why there is so much normative disagreement despite the objective normative facts (Brink, 1984, § 3; Shafer-Landau, 1994), and they have also disagreed about whether normative vagueness should be explained on semantic, epistemic, or metaphysical grounds.[2] This paper, however, focuses on the less discussed question of how metaethical anti-realists could make sense of the instances of alleged normative vagueness.[3] It approaches this question from the direction of Timothy Williamson's (2000, § 4.3) anti-luminosity argument. Let me explain why.

Many philosophers have assumed that we have a privileged access to our phenomenal experiences.[4] These experiences, which it is something like to have, have been assumed to be directly and error-freely accessible, and hence they have been called "transparent" or "luminous". The claim has thus been that, whenever we have a phenomenal experience (when we, for example, feel cold), we are always able to know that we are in that condition. Philosophers have then used the phenomenal realm to build metaphysical systems in which there is no room for radical scepticism.[5]

Williamson's anti-luminosity argument (hereafter the "ALA") challenges the previous dogma. It argues that there are unknowable facts even about our own phenomenal experiences. Yet, as § 2 below explains, the ingredients of the ALA are relatively sparse. The original ALA requires only (i) a Sorites sequence of phenomenal experiences, (ii) knowledge in some stages of that sequence, and (iii) certain general assumptions about knowledge and our doxastic dispositions. The ALA thus seems applicable more generally to any putatively luminous domain in which the conditions corresponding to (i)-(iii) are satisfied.[6]

In ethics, it is often argued that, if an action has some normative property (it is right, good, what we ought to do, and so on), it must be knowable that the action has that property. It is thus argued that, because normative facts are action-guiding and tied to our reactions of praise and blame, they must be knowable to all ordinary agents and hence "luminous".[7] Yet, conditions corresponding to (i)-(iii) above seem

---

[1] See Dougherty, 2018.

[2] See Dougherty, 2018, 186–187. For objections to the semantic account, see Constantinescu, 2014, § 3.1.1, Dougherty, 2014, § 3, and Schoenfield, 2016, § 4.A. For a defence of the epistemic account, see Hawthorne, 2022, and for objections Sider, 1995, Dougherty, 2014, § 2, Constantinescu, 2014, § 3.2.3, and Schoenfield, 2016, § 4.B. For a defence of the metaphysical account, see Schoenfield, 2016, and for objections Constantinescu, 2014, § 3.1.2.

[3] Dougherty (2018, p. 192) recognises this as an area of future research.

[4] For a critical discussion, see Schwitzgebel, 2011.

[5] See, e.g., Ayer, 1936. The phenomenal realm has also been used to understand, for example, rationality, evidence, and the sameness of meaning (Srinivasan, 2015a, p. 294).

[6] See Srinivasan, 2015a, pp. 294–296 and Williamson, 2000, pp. 108–109.

[7] See, e.g., Constantinescu, 2014, pp. 176–182, Schoenfield, 2016, p. 279, Cosker-Rowland, 2017, § 1, and all the references in Srinivasan, 2015b, pp. 279–282 and p. 291 n. 25. Whilst it may be true that normative facts generally must be action-guiding, maybe not every one must be (McGrath, 2010, pp. 73–74; Srinivasan, 2015b, p. 281), and there can also be indirect ways in which the unknowable normative truths

to be satisfied in the normative domain too. There are normative Sorites sequences, we make claims to knowledge within them, and normative knowledge is not assumed to be different in kind. If the ALA is sound, it should therefore also show that there must be unknowable normative facts too as § 2 below will suggest.

The main aim of this paper is to explain the different ways in which metaethical anti-realists can account for the existence of these unknowable normative facts. I believe that there are two good reasons to attempt to provide such explanations. Firstly, one "popular way of defining Realism is as the doctrine that verification-transcendent truth is possible (Wright, 1997, p. 24)." The most explicit defender of this epistemic understanding of the realism versus anti-realism distinction is Crispin Wright (see, e.g., 1981, especially pp. 49–50), but similar views have also been expressed by Dummett (1982, especially p. 55) and Putnam (1981, ch. 3). John Wright (1997, p. 24) summarizes the reasoning that has led many in this tradition to identify realism with the possibility of unknowable facts in the following way:

> Intuitively, Realism is the doctrine that the world exists, and has its properties *independently* of the human mind. If this is so, then the world has its properties independently of human *knowledge* or *belief*, or of the evidence available to us…. A plausible re-expression of this idea is that *the truth about world* is indendent of the evidence available for us… A natural way of clarifying this is to say that there may be truths about the world we cannot know, or for which we cannot have sufficient evidence… So, it seems that maybe a good way of explicating Realism… is to say that there may be truths we can never have sufficient evidence to believe, or can never verify.

However, if realism about a given domain consisted of the possibility of verification-transcendent truths in this way and we were able to use the ALA to show that there are at least some unknowable truths in that domain, these two premises would seem to provide a powerful argument for realism about the domain in question. This is, in fact, an argument Williamson (1996, pp. 908–909) makes. He argues that, in any domain to which the ALA applies, the discourse about that domain "automatically has the potential for truth to outrun what participants can be warranted in asserting (Williamson, 1996, p. 909)," and so realism should be our default stance about the domain in question.

Thus, one goal of this paper is to show that this type of arguments for realism will fail at least in metaethics. It does this by arguing that many of the paradigm forms of metaethical anti-realism allow us to explain the possibility of verification-transcendent normative truths, and therefore such truths cannot be used in an argument against metaethical anti-realism. This is because the claim that unknowable facts are constitutive of realism, a suggestion made by Wright, Dummett, and Putnam (and also accepted by Williamson (1996, p. 907), just is not tenable in the metaethical context as illustrated by this paper's anti-realist explanations of the unknowable normative truths.

---

are relevant for action (Hawthorne, 2022, pp. 232). For a positive defence of unknowable obligations, see Sorensen, 1995 and for a response Sider, 1995.

There is, however, also a less confrontational motivation for the paper. One important project in recent metaethics has come to be known as "quasi-realism". According to Blackburn (1993, p. 4), a quasi-realist "starting from an antirealist position finds himself progressively able to mimic the thoughts and practices supposedly definitive of realism." Within this project, the metaethical anti-realists have made genuine progress by constructing anti-realist accounts of the realist-seeming ways in which we talk about objective normative facts and normative knowledge, make inferences, and so on. Yet, so far, the anti-realists have not attempted to explain how we can also talk about the limits of our knowledge in the context of Sorites sequences, that is, about the alleged cases of normative vagueness. Thus, the paper's second aim is not so much to respond to any potential objections to metaethical anti-realism, but rather to push the quasi-realist project further by providing these explanations in a way that it is hopefully illuminating.

Before we begin, I also want to make one further clarification about the connection between anti-luminosity and vagueness. Even if both are illustrated with the same Sorites sequences, we need to keep in mind that they are distinct phenomena. Even if there are Sorites sequences in which some facts about permissibility are unknowable, this in itself does not entail necessarily that "permissible" is a vague predicate. This is because the unknowable facts of the ALA only indicate the existence of vagueness if the epistemicists are right about what constitutes vagueness. However, if vagueness is a semantic or a metaphysical phenomenon, then the discussion below will merely show that there will be some unknowable normative facts in the relevant Sorites sequences and not that there is also genuine normative vagueness. This is why this paper refers to the normative Sorites sequences merely as *alleged* cases of normative vagueness. Yet, even if this paper focuses on providing anti-realist explanations of anti-luminosity in the relevant Sorites sequences, I will also briefly indicate how the anti-realists could, in addition, accommodate semantic and metaphysical vagueness too (see footnotes 26 and 27).

Here is how the paper will proceed. § 2 first outlines the ALA in the normative domain and responds to three objections. § 3 then explores how metaethical anti-realists (error theorists (§ 3.1), subjectivists, realists, and contextualists (§ 3.2), expressivists (§ 3.3), and the defenders of the response dependence views and constructivism (§ 3.4)) could react to the argument. I will argue that, with two exceptions, the anti-realists need not challenge the ALA, but rather their metaethical views can be used to explain its key consequence, the putative unknowable normative facts. Furthermore, whist reaching this conclusion, we also learn much about how various anti-realist metaethical views can make sense of the alleged cases of normative vagueness more generally.

2 The anti-luminosity argument

This section outlines an ALA in the normative domain.[8] For this purpose, I borrow a normative Sorites sequence from Miriam Schoenfield (2016, p. 262):[9]

> DIVERSIONS: Darryl is watching his two-year-old daughter play in a city park. It is permissible to divert his attention from her for one second. It is not permissible to divert his attention from her for five minutes. It is permissible to divert his attention for 30 seconds? 31? 32? Plausibly, we can create a Sorites series, admitting of borderline cases of permissibility, out of a series of diversions whose lengths differ by millisecond.

This case provides the following premises:

> BEGINNING: Darryl is permitted to divert his attention from her daughter for one second.
> END: Darryl is not permitted to divert his attention from her daughter for 300 seconds (i.e., for five minutes).

For the third premise, we need to assume, for the argument's sake, that all permissibility facts are at least in principle knowable by us. This gives us:

> KNOWABILITY: When an action is permissible, we are in a position to know that the action is permissible.

The fourth premise is Williamson's (2000, p. 97) "margin for error" principle, which is not intuitively obvious, and which also is the most controversial premise of the ALA (see fn. 10 below). We can formulate this premise as follows:

> MARGIN-FOR-ERROR: If we know that it is permissible for Darryl to divert his attention for n seconds, then it is permissible for Darryl to divert his attention for $n+1$ seconds.

The crux of the ALA is that BEGINNING, END, KNOWABILITY, and MARGIN-FOR-ERROR are jointly inconsistent. By BEGINNING, Darryl is permitted to divert his attention from his daughter for one second. By KNOWABILITY, we are in a position to know this. By MARGIN-FOR-ERROR, if we know the previous fact, it is permissible for Darryl to divert his attention for two seconds. By KNOWABILITY, we must again be able to know this second fact, and then by MARGIN-FOR-ERROR the three second distraction too

---

[8] This presentation is inspired by Williamson (2000, ch. 4), the critics of the argument (see fn. 10 below), and especially by Srinivasan (2015a). For previous applications to the normative domain, see Wedgwood, 2010a; McGrath, 2010, p. 72, and Constantinescu, 2014, pp. 175–176. In the normative domain, Hawthorne (2022, § 2.2) also puts forward a related argument that only relies on BEGINNING, END, and standard inference rules for the quantifiers and logical connectives.

[9] For other normative Sorites sequences, see Dougherty, 2014, pp. 353–354, Schoenfield, 2016, pp. 262–263, Dunaway, 2017, pp. 39–41, and Dougherty, 2018, pp. 185–186.

must be permissible. We can then repeat these steps until we reach the conclusion that it is permissible for Darryl to divert his attention from his daughter for five minutes. Yet, by END, a five-minute distraction is not permissible and so we have a contradiction. One of the premises must thus go.

It does not seem like we could reject either BEGINNING or END as they seem to follow from the case itself. We thus seem forced to reject either KNOWABILITY or MARGIN-FOR-ERROR. The critics of the ALA (the "luminists") would reject MARGIN-FOR-ERROR. The problem, however, is that MARGIN-FOR-ERROR seems to follow from an intuitive general safety requirement on knowledge and a plausible empirical hypothesis of our actual doxastic dispositions (Srinivasan, 2015a, § 3).

The relevant safety requirement is the following (see Srinivasan, 2015a, p. 299):

> SAFETY: We know that it is permissible to do a given action only if in all sufficiently similar cases in which we believe that a similar action is permissible it is true that the action in question is permissible.

This principle states that knowing that some action is permissible requires not just a true belief, but also having nearby true beliefs, which seems intuitive.[10] If I happen to be correct in believing that a certain lie is permissible but get the permissibility of other similar lies wrong, my true belief is merely a lucky accident and hence not something to be called knowledge.

SAFETY alone will not, however, entail MARGIN-FOR-ERROR because you might stop believing that Darryl is permitted to be distracted for a certain duration in the sequence exactly where the distraction becomes impermissible (Berker, 2008, p. 8). So, we need a bridge premise: an empirical premise about the doxastic dispositions we have as the kind of creatures we happen to be. This premise can be formulated as follows (Srinivasan, 2015a, pp. 303–304):

> DOXASTIC DISPOSITION: If we believe that it is permissible for S to φ in C, then for any agent S', action φ' and circumstances C' very similar to S, φing and C, we have some disposition to believe that it is permissible for S' to φ' in C'.

This empirical hypothesis seems plausible too. It captures the idea that, in normative thinking, we are disposed to treat like cases alike. If we think that it is permissible for Ann to tell a certain white lie, we are inclined to believe that Ben does nothing wrong if he tells a similar white lie himself.

From DOXASTIC DISPOSITION, we can infer the following, more specific premise (Srinivasan, 2015a, pp. 302–303):

---

[10] Some critics of the ALA reject this principle altogether (see, e.g., Bruecker & Fiocco, 2002, and Neta & Rohrbaugh, 2004). Others claim that either MARGIN-FOR-ERROR requires either an implausible version of the principle or a plausible one that can only be used in the argument with other implausible Soritical assumptions (see, e.g., Weatherson, 2004, Berker, 2008, Ramachandran, 2009, and Vogel, 2010). For a defence of the principle, see Srinivasan, 2015a, and for further objections Vanrie, 2020. For a discussion of what counts as sufficiently nearby, see Srinivasan, 2015a, p. 299 fn. 10.

BELIEF: If we believe that it is permissible for Darryl to withdraw his attention for n seconds, then there exists a sufficiently similar possible case in which someone diverts their attention for $n+1$ seconds and in which we believe that it is permissible for them to do so.

SAFETY and BELIEF then together entail MARGIN-FOR-ERROR (Srinivasan, 2015a, p. 303). If we know that it is permissible for Darryl to withhold his attention for n seconds, then, by SAFETY, there cannot be any nearby cases in which believe that it permissible to withdraw attention for similar durations, but this belief is false. According to BELIEF, it we believe that it is permissible for Darryl to withdraw his attention for n seconds, there are nearby possibilities in which we believe that it is permissible for someone like Darryl to divert their attention for $n+1$ seconds. This means that if we, as a matter of fact, know that it is permissible for Darryl to withhold his attention for n seconds, then it must be permissible for Darryl to also withhold his attention for $n+1$ seconds or otherwise SAFETY would be violated. And, so, we seem to get MARGIN-FOR-ERROR from SAFETY and BELIEF. Hence, insofar as SAFETY is a compelling general condition on knowledge and BELIEF a plausible empirical hypothesis about our doxastic dispositions, we have at least some reason to accept MARGIN-FOR-ERROR. The best way to avoid the contradiction described above then seems to be to give up KNOWABILITY and grant that there must be at least some unknowable normative facts.

There are, of course, several general objections to the ALA that have been discussed in the case of the phenomenal experiences.[11] However, my aim here is not to vindicate the ALA generally, but rather to consider what consequences it would have for the anti-realist metaethical theories if it were defensible. These consequences can be explored without a full vindication of the argument. Rach Cosker-Rowland (2017, pp. 826–827, n. 23) has, however, made two more specific objections to any attempts to apply the ALA to the normative domain. I will address these concerns next together with a third additional concern about DIVERSIONS that allows us to connect the discussion more closely to the existing literature on moral knowledge.

Cosker-Rowland (2017, p. 826 n. 23) first observes that idealized normative reasoners would be able to know all non-normative facts in DIVERSIONS. For every length of time to the millisecond, they would know all the different potential consequences of being distracted for that duration, how probable those consequences would be, and so on. For them thus, somewhere in the sequence there could be a precise point where they would judge that it becomes impermissible for Darryl to divert his attention for that long because the relevant normative threshold for permissible risks would be exceeded at this point. Therefore, arguably BELIEF would not hold for the idealized reasoners, and so it could not be used to derive MARGIN-FOR-ERROR.

Yet, this objection only establishes that normative facts in the Sorites sequences could in principle be knowable for the idealized normative reasoners, who are very different from us, the actual agents. Unlike them, we do not know all empirical facts

---

[11] In addition to the debates about MARGIN-FOR-ERROR, another key debate is about whether the argument relies on a form of Sorites reasoning that can be avoided by the standard views of vagueness (see Williamson, 2000, § 4.5 and Vogel, 2010, part I).

concerning these sequences and we lack the finer discriminatory abilities required for being able to discern to an exact millisecond when the risks become too high[12]. However, as Srinivasan (2015a, p. 317) puts it, to think that the debate is about possible, more able creatures is "to mistake the dialectical situation". The ALA is, after all, compatible with all kinds of creatures who would be much better at discerning which actions are permissible. Instead, the debate really is about whether normative facts are always knowable for the kind of beings we actually are, and Cosker-Rowland's first objection does not challenge the ALA in this respect.

Cosker-Rowland's second objection draws from Wedgwood (2010b, p. 222). It argues that safety-based arguments against normative knowledge generally cannot succeed because normative truths are necessary. To quote Cosker-Rowland (2017, p. 826, n. 23):

> Wedgwood plausibly claims that in order for a safety-based argument against S having knowledge of the moral proposition p to succeed it must be the case that there is a possible world in which p is false. And if moral truths are necessary, there is no such world. So, safety-based arguments against S having knowledge of the moral proposition p cannot succeed.

When making the previous argument, Wedgwood (2010b, p. 222) was addressing a specific version of normative scepticism. Let's assume that my normative beliefs are the result of an ordinary upbring process, and my culture and character. The question Wedgwood asks is whether, given the safety requirement on knowledge, the status of my normative beliefs as knowledge could be undermined by the possibility of a "moral evil demon". He then answers this question "no" for two reasons. Firstly, if my current normative beliefs are true, the demon could not make those beliefs false because the normative facts that make my beliefs true are necessary and thus obtain in all worlds. Secondly, the possible scenario in which the demon would give me different moral beliefs is arguably not a nearby scenario that would be relevant for the application of the safety requirement. On these grounds, Wedgwood (2010b, p. 222) concludes that the moral evil demon cannot provide "a compelling reason to doubt the safety of our moral beliefs".

Yet, pace Cosker-Rowland, the previous response to the moral evil demon challenge does not entail that SAFETY cannot be used in the ALA to support MARGIN-FOR-ERROR. Let's assume that we know that it is permissible for Darryl to divert his attention for n seconds. The ALA does not rely on possible worlds in which that same belief would be false (as none exist) to argue for MARGIN-FOR-ERROR. Rather, it only

---

[12] It is true that the debate about luminosity is about whether we are always "in a position to know" normative facts and not about whether we, as a matter of fact, know them. This could be understood as a question of whether we would know those facts under some idealized hypothetical circumstances in which had more evidence and reasoned better. However, here the idealization is more modest– we still assumed to have our actual ordinary cognitive capabilities with merely some modest improvements and some relatively easily attainable additional knowledge, the kind of improvements we could realistically undergo in real life. In contrast, the idealized reasoners considered in Cosker-Rowland's objection would need to be idealized in much more radical ways, which is why they are not really germane to the issue of our knowledge in the ALA.

relies on similar and yet marginally different nearby cases in which Darryl would divert his attention for slightly longer durations. It is true that, for any such longer duration, it is either necessary or impossible that it is permissible for Darryl to divert his attention for that long. Yet, all the argument requires given its SAFETY premise is that, if we know that it is permissible for Darryl to divert his attention for n seconds and we also simultaneously believe that it is permissible for him to also divert his attention for $n+1$ seconds, then that belief must also be true. Nothing about normative truths being necessary truths challenges this requirement. Cosker-Rowland's two concerns should thus not prevent us from examining the ALA further in the normative domain.

I want to address also a third concern one might have about DIVERSIONS. Mark Schroeder (2014, § 6.2) has introduced a distinction between mixed and pure normative claims. Mixed normative claims are in part about contingent empirical facts and in part about foundational normative principles, whereas the pure ones are solely about those principles. It could then be observed that, in DIVERSIONS, the claim that it is permissible for Darryl to divert his attention for n seconds really is a mixed normative claim. It is in part a claim about what kind of consequences that action would have and in part a claim about some foundational normative principle that determines whether bringing about those consequences is permissible. This leads to the objection that the reason why we cannot know whether the mixed claim in question is true is that we just cannot know the empirical consequences of the n-second distraction and so we cannot know how to apply the relevant foundational normative principle to the case. This in itself, however, seems to give us no reason to doubt our ability to know whether the foundational normative principle is true, which could be argued to be the main concern of the existing literature on moral knowledge.[13]

In response, it should be noted that we seem able to apply the ALA also to the foundational normative principles themselves. Take the utilitarian principle according to which we ought to maximize the total amount of happiness.[14] Let us also assume that Daniel Haybron's (2008) popular account of happiness is along the right lines: happiness consists of a person's emotional condition as a whole, of (i) endorsement states such as joy, (ii) engagement states such as flow, (iii) attunement states such as tranquility, and (iv) dispositions to experience various positive moods. This view entails that there are various 'happiness-like' properties, $happiness_1$, $happiness_2,\ldots$, $happiness_n$ that all have a slightly different extension. To be $happy_1$ might require having all the relevant emotional states in a strong form and all the mood dispositions, whereas at the other end of the Sorites sequence to be $happy_n$ might only require a single very narrow mood disposition.

In this situation, we can use the ALA to argue that there are unknowable truths also about the previous foundational normative principle itself. If there are two actions available for an agent between which the only difference is that the first action makes someone else $happy_1$ whereas second leaves that person in an emotionally neutral state, we know that the foundational principle in question requires the agent to do the

first action. By contrast, if the only difference between the actions were whether the other person would be happy$_n$ or not (where this is a question of just one insignificant mood propensity), we know that the foundational principle would be neutral between the actions. We can then use the ALA in the way outlined above to argue that there are some unknowable facts here about the foundational normative principle itself. For some of the happiness-like properties in the Sorites sequence we will not be able to know whether the foundational utilitarian principle itself requires us to promote them or not. Furthermore, in this case, the unknowable normative facts are not due to the empirical elements of mixed normative claims because here we, by stipulation, know all the facts about what consequences our actions have for the emotional states and mood dispositions of others. This argument suggests that the ALA also applies to the kind of foundational normative principles on which the existing literature on moral knowledge tends to focus.

3 Anti-realist responses

If the ALA applied to the normative domain in the way just outlined, it would establish that there must be at least some unknowable normative facts. If you are a metaethical realist who thinks that normative facts are a part of the stance independent fabric of the universe itself, it is natural to think that there can be such facts we can never come to know. The question, however, is whether the anti-realists too can explain how such unknowable moral facts could obtain. Are there ways for the anti-realists too to accommodate these facts in the normative Sorites sequences?

The rest of this paper tries to show that, with few exceptions, the metaethical anti-realists can accommodate these unknowable normative facts in their metaethical frameworks too. Thus, the following sub-sections consider how error theorists (§ 3.1), subjectivists, realists, and contextualists (§ 3.2), expressivists (§ 3.3), and the defenders of the so-called response dependence views and constructivists (§ 3.4) could react to the ALA. They argue that all other anti-realists except the error theorists and the defenders of certain versions of the response dependence views and constructivism can accommodate the argument's key implications, the unknowable normative facts, and so the ALA cannot be used to motivate metaethical realism. Furthermore, in addition, we will learn much about how metaethical anti-realists can makes sense of the alleged cases of normative vagueness.

3.1 Error theory

According to metaethical error theories, normative sentences are truth-apt but never true.[15] Error theorists then typically motivate this view with an argument that has two premises. According to the first, semantic premise, normative discourse presupposes (as a part of the meaning of normative terms) "non-negotiably" that, if any normative facts were to obtain, they would have to have some feature F such as inherent "to-be-doneness" or categorical authority. According to the second, metaphysical premise,

---

[15] See, e.g., Mackie, 1977, Joyce, 2001, and Olson, 2014.

the feature F is fundamentally problematic and so no facts with the feature F obtain. From these two premises, many error theorists conclude that there are no normative facts, normative properties are not instantiated, and thus all positive normative claims are false.

In response to the ALA, the error theorists would reject its first two premises: BEGINNING and END. They would argue that both being permissible and being impermissible are normative properties that no action can have.[16] They would thus deny the existence of normative Sorites sequences as such sequences would require the instantiation of normative properties, which the error theorists reject.[17] This means that the ALA cannot be used as an argument against the error theory in a non-question-begging way as putting forward the first premises of the argument would require assuming the intended conclusion.

Furthermore, having the following two commitments is arguably required for being a realist about some domain D (Devitt, 1983, pp. 75–76):

EXISTENCE: The objects and properties in D exist.
STANCE INDEPENDENCE: The objects and properties in D do not fundamentally depend on human attitudes, judgments, beliefs, or the like.

The problem for the normative ALA is that such an argument could at best support STANCE INDEPENDENCE, but it could not be used to justify EXISTENCE as two premises of the argument would presuppose that thesis. The ALA was never, however, intended to be an argument against error theories in the first place. According to its defenders, if an error theory is true of some property, the condition of that property being not instantiated is trivially luminous because we could always know whether that condition obtains: "A condition that obtains in no case, the impossible condition, is automatically luminous" (Williamson, 2000, p. 107).

3.2 Subjectivism, relativism, and contextualism

There is a family of subjectivist, relativist, and contextualist views according to which (i) there is no objectively correct normative standard and (ii) the truth of normative claims is in some way grounded in the approvals and disapprovals of an individual or a community.[18] These views thus reject STANCE INDEPENDENCE, which makes them forms of anti-realism. This sub-section first introduces a traditional version of simple subjectivism and explains how it can respond to the ALA. It then suggests that this response can also be given by the more sophisticated subjectivists, relativists, and contextualists.

According to traditional simple appraiser subjectivism, if I use the normative predicate "… is permissible" to make the assertion that eating meat is permissible, my assertion means that I have an attitude of approval towards eating meat (see

---

[16] This is consistent because the error theorist can deny both that "not permissible" entails "impermissible" and that "not impermissible" entails "permissible" (Olson, 2014, p. 14).

[17] Error theorists would thus deny that normative vagueness exists (Dougherty, 2018, p. 191).

[18] For surveys, see Rachels, 1991, Wong, 1991, Dreier, 2009, and Francén Olinder, 2016.

Rachels, 1991, p. 435). Given the meaning of the predicate, my assertion reports that I have a positive attitude towards such actions. According to this view thus the referents of normative predicates and the contents, truth-makers, and truth-conditions of normative claims are all given by the speaker's own approvals and disapprovals.

According to the previous view, you can assert truthfully that it is permissible for Darryl to divert his attention from his daughter for n seconds if and only if you have an attitude of approval towards Darryl doing so. DIVERSIONS is, however, also a Sorites sequence concerning your own attitudes of approval and disapproval. When you consider Darryl diverting his attention for just one second, you are likely to have a strong attitude of approval towards him doing so. As the durations get longer, your approval presumably gets weaker and gradually turns into weak disapproval first and finally to stronger disapproval.

Approvals and disapprovals are, however, in many ways like phenomenal experiences. This allows us to apply the ALA also to the previous sequence of attitudes in DIVERSIONS. In the beginning of the sequence, you know that you approve of Darryl diverting his attention for just one second. You must then also approve of Darryl diverting his attention for two seconds or the previous belief would not satisfy the safety requirement on knowledge and hence the margin for error principle it supports together with our doxastic dispositions (see § 2 above). If we then assumed that approvals were luminous states and so you would know that you have that attitude of approval also in the two second case, then you must also have it towards a three-second distraction, and so forth. However, if the ALA is sound and we want to avoid the contradiction in the end, there must be a step in the sequence where you know that you have an approving attitude towards a distraction of n seconds and you have an attitude of approval towards a $n+1$-second distraction even if you cannot know that.[19] If this weren't the case, we would get the absurd conclusion that you have an attitude of approval towards Darryl diverting his attention for 5 minutes. In this Sorites sequence thus, there must be some durations for which you cannot know whether you approve of Darryl diverting his attention for that long even if there is a fact of the matter.

Recall then that, according to simple subjectivism, you can truthfully assert that it is permissible for Darryl to divert his attention for n seconds if and only if you, as the appraiser, have an attitude of approval towards him doing so. This entails that, for you to be able know that it is permissible for Darryl to divert his attention for n seconds, it presumably is necessary that you know whether you approve of him being distracted for that long. But, given the previous ALA, for some durations there are facts about your own disapproval that you cannot know. In these cases, simple subjectivism entails that you also cannot know whether the action in question is permissible.[20]

This means that simple subjectivism and the ALA concerning our own attitudes together entail that there will be some permissibility facts in DIVERSIONS that we

---

[19] Here we need to assume that having the weakest possible disapprovals and no disapproval at all are at least close to being phenomenally indistinguishable given our powers of discrimination. This would make a hypothesis corresponding to DOXASTIC DISPOSITION plausible here too.

[20] This is also pointed out by Srinivasan (2015b, p. 277).

cannot know, assuming that subjectivism is true. These are, however, the same kind of unknowable normative facts that the ALA concerning the normative properties themselves appeared to support in § 2, and so the ALA cannot be used as an argument against simple subjectivism. The ALA can thus at best establish that there are unknowable normative truths within the normative Sorites sequences, but it cannot show that these truths are stance independent. This is because the ALA itself supports corresponding unknowable facts about our own attitudes that could in principle ground those unknowable normative truths.

Similar claims can be made also about more sophisticated forms of subjectivism and relativism. More complex subjectivist views ground normative properties in the hypothetical attitudes an appraiser would have if she were better informed and more coherent, imaginative, rational, and/or sensitive.[21] Similarly, relativist meta-ethical views ground normative facts in the attitudes of a community.[22] In both cases, however, the attitudes of the idealized version of the agent and the attitudes of a community form the same kind of a Sorites sequence as in the case of the simple subjectivism above. This means that, for these theories too, the sophisticated subjectivists and relativists can use the ALA to show that there must be some unknowable approval facts about the attitudes of the idealized appraiser and the attitudes of the community in the normative Sorites sequence. And, in the context of these views, these facts would then ground the corresponding unknowable normative permissibility facts that were the conclusion of the normative ALA.

Contextualist views work in a slightly different way, but they have similar consequences. According to these views, our normative assertions do not report our attitudes, but rather in the context of an utterance those attitudes fix the reference of normative predicates to specific natural properties.[23] The way in which this reference-fixing happens is determined by the "character" of the normative predicates, a rule that determines how a certain property gets assigned to be a given predicate's semantic value in the context of utterance.

There are, however, many conceptions of properties according to which properties are abundant.[24] According to the maximalist versions of these views, there is a property that corresponds to any set of objects whatsoever. Let us then consider DIVERSIONS and the properties that could be the semantic value of the normative predicate "permissible" in it. Here, for property P_1, only Darryl's action of withdrawing his attention for one second would fall under its extension, for P_2, both one and two second distractions would fall under its extension, and so on until the property P_{500} for which all distractions between one and 500 seconds would fall under its extension (Dunaway, 2017, p. 43).

In the contextualist framework, our approval attitudes determine to which of these properties the reference of "permissible" gets fixed in DIVERSIONS. We already saw, however, that in the middle of the sequence there are some durations for which we cannot know whether we have these attitudes or not. Because of this, it is hard to see

[21] See, e.g., Williams, 1981, Darwall, 1983, part 2, and Sobel, 2009.

[22] See, e.g., Wong, 1984, Harman, 1995, and Velleman, 2015.

[23] Such views were first introduced by Dreier (1990). For a survey of recent developments, see Silk, 2018.

[24] See, e.g., Lewis, 1983.

how we could know to which one of the previous abundant properties our attitudes would fix the reference of "permissible" even if there were a fact of the matter. Therefore, due to the unknowability of our own reference-fixing attitudes, contextualism too appears to create unknowable normative facts in the Sorites sequence.[25]

Overall, then, different forms of subjectivism, relativism, and contextualism are perfectly compatible with the ALA. Their defenders can argue that their metaethical views can explain the obtaining of the unknowable normative facts in the normative Sorites sequences.[26] Because of this, the explanatory challenge created by the ALA does not seem to give us any reason to favour metaethical realism over anti-realism. Yet, we have also learned much about how the subjectivists, relativists, and contextualists can understand the alleged cases of normative vagueness.

3.3 Expressivism

There are many versions of expressivism, but for the sake of simplicity I will focus on a simple formulation by Schroeder (2008). I will also draw from Simon Blackburn's (1984 and 1993) expressivist views of normative truth and knowledge.

According to Schroeder's (2008, § 4.2) expressivism, for each normative predicate, given its meaning, there is a certain attitude corresponding to it. So, for example, "permissible" perhaps stands for approval, "wrong" for blaming, "better than" for preferring, and so on. Let us then take the sentence "It is permissible for Darryl to divert his attention for n seconds". According to this view, its meaning consists of the fact that it conventionally expresses the general non-cognitive, positive attitude of *being for* having the attitude of approval towards Darryl diverting his attention for n seconds. That attitude of *being for* here corresponds to believing, and the attitude of approval to the content of this belief.

We can then apply Blackburn's (1984, p. 198) original attitudinal account of normative truth to the previous formulation of expressivism. If we do this, we get a view according to which the content of your "belief" that it is permissible for Darryl to divert his attention for n seconds is true if and only if the attitude of approval towards being distracted for n seconds (towards which you have the attitude of being for, i.e., the believing part) is an attitude that would belong to the best possible set of attitudes resulting "from taking all possible opportunities for improvement of attitude" such as making them more coherent and consistent (ibid.). Finally, we also need an expressivism-friendly account of normative knowledge. According to Blackburn

---

[25] Alternatively, the reference-fixing attitudes might leave it open to which property the reference gets fixed, which would be a source of semantic indeterminacy (see Dougherty, 2014, pp. 365–6).

[26] These views thus seem compatible with epistemic accounts of vagueness, but such accounts arguably require that the relevant ignorance is due to semantic plasticity, which may not hold here. Hawthorne (2022, p. 220) thus suggests that Diversions is a case of pure ignorance rather than vagueness. For a discussion of whether epistemicists need to explain the relevant ignorance with semantic plasticity, see Schoenfield, 2016, pp. 277–281 and the references therein. Subjectivism, relativism, and contextualism also seem compatible with semantic accounts (if it is not semantically determined exactly how strongly you must approve of an action to be able to truthfully say that it is permissible) and metaphysical accounts (insofar as we can have vague attitudes (see Dougherty, 2014, p. 362). There could also be many simultanous sources of vagueness (Dunaway,2017, p. 41). Also, that these views can accommodate unknowable normative facts may undermine the motivation for them (Srinivasan, 2015b, pp. 276–277).

(1993, p. 90), "[t]o know something is to know that no judgment contradicting one's own could be really preferable" or, in other words, it is to know that "no improved standpoint, yielding a revised estimate, is possible". Knowing that Darryl is permitted to divert his attention for n seconds would thus consist of knowing that the attitude of approval towards n-second distractions belongs to the best possible set of attitudes you could have.

We can then return to DIVERSIONS and the ALA. The conclusion of the ALA was that there must be a step in the sequence where (i) we know that it is permissible for Darryl to divert his attention for n seconds, (ii) we truly believe that it is permissible for him to divert his attention for $n+1$ seconds, and yet (iii) we cannot know that to be the case. Assuming expressivism, would it make sense to claim that such cases must exist?

As we saw, according to expressivism, knowing that it is permissible for Darryl to divert his attention for n seconds consists of knowing that approval of an n-second distraction belongs to the set of attitudes you would have if your attitudes were maximally improved. Presumably for all the shortest durations in DIVERSIONS we know that our best attitudes would include approval. After all, the very short distractions do not impose any risk on Darryl's daughter or anyone else, and so we get the required instances of (i).

What about (ii) and (iii)? For you to believe truly that it is permissible for Darryl to divert his attention for $n+1$ seconds requires, according to expressivism, both (a) that you are for having an attitude of approval towards a $n+1$-second distraction and (b) that your best set of attitudes too would include that same attitude of approval. Furthermore, for you not to be able to know that it is permissible for Darryl to divert his attention for $n+1$ seconds (as you truly believe) requires that you cannot know that your best possible set of attitudes would still contain an attitude of approval towards $n+1$-second distractions (even when the best set does contain that attitude).

Would it then make sense for you to be in this epistemic state somewhere in the sequence? Both given BELIEF and intuitively, if you know that the attitude of approval towards a distraction of n seconds belongs to the best possible set of attitudes you could have, then surely you are disposed to be for having an attitude of approval also towards a distraction of $n+1$ seconds. And, given that your ideal set of attitudes contains an attitude of approval towards a distraction of n seconds, that best set really can also include an attitude of approval towards a distraction of $n+1$ seconds. According to expressivism then, this is all that is required for you to truly believe that the distraction of $n+1$ seconds is permissible, and so we get (ii) too.

It is not, however, immediately obvious that, at this point of the sequence, you would be anymore able to know that the attitude of approval towards the $n+1$-second distraction belongs to your best possible set of attitudes (even when it does). This is because, as the durations get gradually longer, at some point you will, for the first time, fall short of being fully reliable in judging whether a given attitude would still belong to your best possible set of attitudes (as the safety requirement on knowledge corresponding to SAFETY would require). After all, from our current imperfect epistemic positions, it is not wholly transparent to us where exactly the boundaries of our approving attitudes would lie in DIVERSIONS if our cares and concerns were made fully coherent, unified, and informed. We thus seem to get (iii) as well.

This means that (given our epistemic situation, our current actual attitudes of approval, our best sets of approval, and the key premises of the ALA), expressivism seems able to make sense of the conditions (i)-(iii) obtaining at some point in DIVERSIONS. Expressivists can thus grant for the defenders of the ALA that, given BEGINNING, END, and MARGIN-FOR-ERROR, there will be a point in DIVERSIONS where you know that a distraction of n seconds is permissible, where you truly believe that $n+1$-second distraction is permissible too, and where that true belief fails to count as knowledge.[27] The expressivists can explain what knowledge and attitudes we have in such cases and why having those attitudes and that knowledge makes sense. This means that the expressivists too seem to be able to explain how the unknowable normative facts can obtain in the Sorites-sequences, and they too can give an account of the alleged cases of normative vagueness.

3.4 Response dependence and constructivist views

The final type of anti-realism I will consider are various forms of response dependence views and constructivism ("rdc-views" hereafter).[28] These views are usually introduced with the following schema:

> BASIC SCHEMA: Necessarily, an action has a normative property N if and only if subject S would in circumstances C would have a reaction R to the action.

This schema is to be read from right to left. That is, the given normative property of an action is, according to the rdc-views, to be explained in terms of the reactions the relevant subjects would have to the action.[29] Because of this, these views are usually presented as forms of anti-realism.

We then get different versions of the rdc-views depending on how we further specify S, R, and C in the BASIC SCHEMA. The defenders of the rdc-views usually begin by first describing the hypothetical idealized subjects S and circumstances C. They may, for instance, claim that the subjects whose reactions matter should be fully rational, fully informed, impartial, imagining vividly, in a reflective equilibrium, have completed certain complex reasoning patterns, and so on.[30] They can also claim that the relevant circumstances must consist, for example, of freedom from domination, bias,

---

[27] Expressivism, more generally and pace Baima (2014, § 3) and Sorensen (1990), seems compatible with the same accounts of the grounds of vagueness as the subjectivist, relativist, and contextualist views (see fn. 26 above) and for the same reasons. Sometimes it is argued that making sense of uncertainty in the borderline cases is a problem for the expressivists (Baima, 2014; Dougherty, 2018, pp. 191–192). For a response, see Sepielli, 2012.

[28] Sometimes these views are also called "Best Opinion" accounts of normative properties (Miller, 2003, ch. 7).

[29] These views are thus modelled upon secondary quality accounts of colours (Wright, 1988). According to these views thus, normative claims are about actions' dispositions to elicit certain reactions in us rather about our own attitudes (see § 3.2 above).

[30] See, e.g., Brower, 1993, Johnston, 1989, and Lewis, 1989. Some defenders of the rdc-views also idealize the relevant subjects and circumstances in normative terms: the subjects must, for example, be fully virtuous, morally good, embracing all morally relevant considerations, or their responses merited (see, e.g., McDowell, 1985). These latter views are usually considered as forms of realism (Wright, 1988).

distractions and the like and enable full attentiveness, focus, open-mindedness, clarity, and stability of judgment.

If the relevant circumstances and subjects are idealized sufficiently in these ways, the rdc-views can respond to the ALA in the same way as the expressivists. Their defenders can argue that we can know that it is permissible for Darryl to divert his attention for one second because we can be confident that the idealized subjects would react positively to such actions given that they pose no risk to anyone. They can then add that, at some point in the sequence, there must be a point where (i) we still know that diverting attention for n seconds is permissible, (ii) we are inclined to believe that a $n+1$-second distraction too is permissible (as per BELIEF), (iii) this belief is also true given that the idealized subject really would still have a positive reaction to the $n+1$-seconds distraction, but (iv) this belief fails to count as knowledge because at this stage our beliefs about the fully idealized subjects' responses are no longer wholly non-accidentally true. It is, after all, difficult to know where exactly the boundaries of their reactions lie. Hence, because many rdc-views idealize the relevant subjects and circumstances sufficiently, the difficulty of knowing where exactly the boundaries of the relevant reactions lie offers us an explanation of the unknowable normative truths of the ALA.

This explanation is not, however, available for all rdc-views. Some versions of these views idealize the relevant subjects and circumstances much more modestly because they place an epistemic constraint for the idealization process. These views begin from the idea that, for the BASIC SCHEMA to capture how the normative properties are "relative to us," the way those properties are grounded in human reactions must be characterised in a way that is easily accessible to us (see Dunaway, 2018, p. 142). According to these views, we can therefore idealize the relevant subjects and circumstances only to the extent that we can still always know what the idealized subjects' responses would be.[31] These modestly idealized versions of the rdc-views cannot therefore use the expressivists' response to the ALA.

We also need to consider the relevant reactions R in the BASIC SCHEMA. The first option is to take these reactions to consist of attitudes such as approval and disapproval. We then get a view according to which an action is, for example, permissible if and only if the modestly idealized (hereafter "normal") subjects in the modestly idealized ("normal") circumstances would approve of the action. When the rdc-views are formulated in this way, their defenders can respond to the ALA in the same way as the subjectivists, relativists, and contextualists (§ 3.2). They can observe that the attitudes of the normal subjects too form a Sorites sequence in DIVERSIONS. Thus, we will know that they have strong approving attitude towards the very short distractions. Yet, when the normal subjects' approving attitudes get weaker, given the ALA itself at some point in the sequence we will only be able to truly believe that they still approve of a given distraction without being able to know that they do. And, given the BASIC SCHEMA, these unknowable attitudes of the normal subjects in the sequence will ground corresponding unknowable normative facts: the same unknowable normative facts the ALA itself insists there must be.

---

[31] Views that tie truth to the final verdicts of idealized and yet epistemically accessible inquiry in this way are considered to be paradigms of anti-realism (see e.g., Putnam, 1981 and Wright, 1992).

The defenders of the rdc-views could, however, also formulate their view in terms of normative judgments. According to the resulting views, an action would be permissible if and only if the normal subjects in the normal circumstances would judge it to be permissible. It could then be argued that normative judgments are binary states: either you judge that an action is permissible, or you don't. Furthermore, intuitively at least, we could worry about whether there really could be unknowable whether someone judges that some action is permissible. It could be suggested that we would always be able to know that, for example, by simply asking them. And, so, it is not clear whether the judgment-based formulations of the modestly idealizing rdc-views could generate the kind of unknowable facts about the relevant judgments that would be needed to ground the corresponding unknowable normative facts.[32] How could the defenders of these views then respond to the ALA?

I want to conclude by suggesting that they too can provide a response that has two components. This response will, however, require both challenging and slightly modifying one of the key premises of the ALA in a way that will be controversial. The first stage of this response requires moving from SAFETY to a weaker principle (inspired by Vogel (2010, p. 549):[33]

> SAFETY*: We know that it is permissible to do a given action only if in all sufficiently similar cases in which we believe that a certain action is permissible it is not false that the action is permissible.

SAFETY and SAFETY* differ only in the cases in which the truth-value of a proposition about an action's permissibility is indeterminate. Imagine a situation in which we believe that some action is permissible and where in some nearby cases where we retain this belief it is neither true nor false that the action in question is permissible even if in no such nearby case it is false that it is permissible. Here, SAFETY entails that our original belief does not count as knowledge whereas according to SAFETY* it could still count as knowledge.

Let us then return to the rdc-views that idealize the relevant subjects and circumstances only modestly. The second stage of the response begins from the observation that, after these idealizations, we are still left with many normal subjects who all have slightly different normative sensitivities. The modest idealizations of the subjects and circumstances thus fail to guarantee full unanimity between the "faultless" normative judgments of the idealized subjects (see McGrath, 2010, p. 60).

In DIVERSIONS, all modestly idealized subjects would probably judge that it is permissible for Darryl to divert his attention for the shortest durations and likewise they would agree that the longest distractions are not permissible. Yet, in the middle of the

---

[32] In response, it could be argued that normative judgments consist of credences that range from full conviction to full uncertainty. This would allow us to construct a Sorites sequence of conviction and use the ALA to show that in this sequence there must be unknowable facts about whether a subject makes the relevant judgment (see Hawthorne (2005). The rdc-views based on normative judgments would then generate corresponding unknowable normative facts in the normative Sorites sequences, and so the rdc-views formulated in terms of normative judgments too could rely on the subjectivists' response to the ALA. I thank the anonymous referee of *Synthese* for this observation.

[33] For the costs of this move, see Srinivasan (2015a, pp. 299–301).

sequence, some normal subjects will faultlessly judge that it is permissible for Darryl to divert his attention for a given duration whereas others will judge equally fault-lessly that it is not. Consider then a version of the rdc-views according to which (i) it is true that an action is permissible if and only if *all* normal subjects in the normal circumstances judge that the action is permissible and (ii) it is false that an action is permissible if and only if *all* normal subjects in the normal circumstances judge that the action is not permissible. This view entails that, in the middle of the Diversions sequence, it is neither true nor false that it is permissible to be distracted for the dura-tion in question (see Shafer-Landau, 1994, p. 337).[34]

Let us then consider a case in which we still judge that it is permissible for Darryl to divert his attention for n seconds. Given that the modestly idealized subjects would all still agree with us at this point, that belief would be true according to the previous view. It is also difficult to question that belief's status as knowledge because we are sufficiently similar to the modestly idealized judges. Now, let us also assume that, due to the doxastic dispositions captured by Belief, we are also inclined to believe that a distraction of $n+1$ seconds would be permissible and so we come to acquire that belief too. Let us, however, assume that $n+1$ is the first case in which the fault-less judgments of the modestly idealized judges diverge, and so according to the view under consideration here it is neither true nor false that the $n+1$-second distraction is permissible.

If we accept the stronger Safety principle and we know that the n-second distrac-tion is permissible, then it must be true that the $n+1$-second distraction is permissible as well or otherwise Margin-for-Error, which is entailed by Safety and Belief, would not be satisfied. This must be the case even if the latter normative fact must remain unknowable to us insofar as we want to avoid the contradiction at the heart of the ALA. If we accept Safety, we must thus reject the rdc-views under consideration. This is because, according to them, normative truths are grounded in the unanimous reactions of the modestly idealized subjects, which are (i) evidentially constrained and hence always knowable and therefore (ii) not able to guarantee that the previous kind of required unknowable positive normative facts would obtain in Diversions.

If, however, we accept only the weaker Safety* principle, we can avoid the pre-vious conclusion. With that principle in mind, even if we were inclined to believe that the $n+1$-second distraction is permissible (given Belief), the belief that the n-second distraction is permissible would still remain sufficiently safe to count as knowledge despite it being not true that the $n+1$-distraction would be permissible. This is because Safety* and Belief do not entail Margin-for-Error but merely the weaker principle:

Margin-for-Error*: If we know that it is permissible for Darryl to divert his attention for n seconds, then it is not false that it is permissible for Darryl to divert his attention for $n+1$ seconds.

---

[34] This view thus creates truth-value gaps, and so the law of the excluded middle will fail in these cases. This furthermore entails that certain laws of classical logic such as the double negation elimination and the modus tollens will no longer be valid. We would thus need to adopt a three-valued system of logic such as the K3 (see Beall & Logan, 2017, part IV and Priest, 2008, chs. 7–9).

In the case under consideration, the belief that it is permissible for Darryl to divert his attention for $n+1$ seconds is neither true nor false because the judgments of the normal subjects diverge for the first time at that point. Our belief that the n-second distraction is permissible will still, however, satisfy the Margin-for-Error* principle and thus that belief can still count as knowledge under that principle. This consequence also stops the *reductio* of the ALA. Even if we were to accept Knowability, Margin-for-Error* would not offer us a route from Beginning to the negation of End. This is because knowledge of the permissibility of a n-second distraction would now no longer require the obtaining of the positive fact that the $n+1$-second distraction too is permissible. Therefore, if the most we can get from a plausible safety requirement on knowledge is Safety* and hence Margin-for-Error*, the ALA fails to pose a problem even for the evidentially constrained versions of anti-realism.

The question then really is: which is more plausible, Safety or Safety*? In this context, from the anti-realist perspective there is some genuine philosophical motivation for preferring Safety* (Wright, 1996, p. 936).[35] The basic thought would be that, if you believe that some action is permissible in a case where even the best opinions of the normal judges in the normal circumstances diverge, the fact that your belief is not true in this case does nothing to suggest that you would be guilty of epistemic shortcomings. Your judgment in this case is, after all, as good as anyone else's. And, because due to no fault of your own you fail to get this particular judgment right as there is no positive fact to be right about here, it is difficult to see why you could not still be making non-accidentally true and thus safe judgments in the previous cases in the sequence in which the judgments of all normal subjects converge.[36] There does not seem to be any reason to think that the divergence of attitudes in the later cases takes anything away from the reliability of your judgment in the previous ones where there is unanimity.

Yet, I acknowledge that this reasoning will not convince everyone. The question then will be: where does it leave the defenders of the rdc-views if they cannot rely on Safety* in the way just outlined? In that case, they would have two options. Either (i) they would have to idealize the relevant subjects and circumstances more so that

---

[35] Manley (2007) argues that, when there is a risk of failure of reference of a singular term, one will lack knowledge even in the case of successful reference as a thought with a failed attempt to refer in the nearby scenario would lack a truth value. This argument seems to count against Safety* in its general form. To avoid this problem, Manley also offers a revised safety condition according to which S knows that p only if S could not easily have had a failed counterpart thought due to, for example, failure of reference. This means that the anti-realists in metaethics who adopted Manley's revision would have to insist that when the idealized subjects disagree about some normative issue the counterpart thought in question does not count as a "failed thought" in the required sense even if it lacks a truth-value.

[36] It could be objected that, in ordinary non-normative cases too, in the nearby scenarios we can come to adopt false beliefs through no fault of our own, but in these cases this possibility genuinely seems to undermine our knowledge. For example, if there is someone stealing cars in the neighborhood, I will not know that my car is parked in front of my house based on my memory of parking there even if not knowing about the thefts is not my fault. Yet, such cases have two important differences to the normative cases above. Firstly, in the normative domain, in the nearby cases it is stipulated that there is nothing more to be right about given that the best judgments diverge. Secondly, in the normative domain, the relevant subject also satisfies the conditions for making the very best judgments that could be made about the subject matter. Neither one of these conditions is satisfied in the car theft example in which the failure of Safety does seem to undermine knowledge.

they could rely on the expressivists' response to the ALA or (ii) they would have to take the relevant responses to consist of attitudes of approval and disapproval so that they could rely on the subjectivists, relativists, and contextualists' response. Otherwise, their only option would be to provide sufficient arguments against SAFETY and for the weaker SAFETY* principle.

4 Conclusion

This paper first outlined the ALA as it would apply to the normative properties. § 2 did not defend this argument against all general objections, but it did address Rach Cosker-Rowland's objections to the idea that the ALA could be applied within the normative domain and also suggested that the argument applies to the foundational normative principles too.

The rest of this paper considered how the defenders of different forms of meta-ethical anti-realism could respond to the ALA. It suggested that (i) the error theorists would object to the required Sorites sequences, (ii) the subjectivists, relativists, and contextualists would ground the required unknowable normative facts in the unknowable attitudes within the relevant Sorites sequences, (iii) the expressivists would explain the unknowable normative facts by highlighting how difficult it would be to know where the boundaries of our idealized attitudes would lie, and (iv) the defenders of the response dependence views and constructivists could either rely on the previous strategies or they could defend a weaker safety requirement on knowledge. As a consequence, it seems that the anti-realists do have satisfactory responses to the explanatory challenge posed by the unknowable normative facts to which the ALA seems to lead. Reaching this conclusion has, however, also helped us to begin to understand how metaethical anti-realists could generally think about the alleged instances of normative vagueness.

Declarations

Competing interests I have no interests that are related directly or indirectly to the work submitted. I do not have any grants by organisations that would gain or lose financially through the publication of the manuscript. I have never been employed by an organisation that would gain or lose financially. And, I do not have any financial interests or non-financial interests in the publication of the manuscript either.

References

Ayer, A. J. (1936). *Language, truth, and logic*. Gollancz.

Baima, N. (2014). The problem of vagueness for expressivism. *Ethical Theory and Moral Practice*, *17*(4), 593–605. https://doi.org/10.1007/s10677-014-9493-0.

Beall, J. C., & Logan, S. A. (2017). *Logic: The basics* (2nd ed.). Routledge.

Berker, S. (2008). Luminosity regained. *Philosophers' Imprint*, *8*(2), 1–22. http://hdl.handle.net/2027/spo.3521354.0008.002.

Blackburn, S. (1984). *Spreading the word*. Clarendon.

Blackburn, S. (1993). *Essays in quasi-realism*. Oxford University Press.

Brink, D. (1984). Moral realism and the sceptical arguments from disagreement and queerness. *Australasian Journal of Philosophy*, *62*(2), 111–125. https://doi.org/10.1080/00048408412341311.

Brower, B. W. (1993). Dispositional ethical realism. *Ethics*, *103*(2), 221–249. https://doi.org/10.1086/293494.

Bruecker, A., & Fiocco, M. O. (2002). Williamson's anti-luminosity argument. *Philosophical Studies*, *110*(3), 285–293. https://doi.org/10.1023/A:1020627215034.

Constantinescu, C. (2014). Moral vagueness: A dilemma for non-naturalism. In R. Shafer-Landau (Ed.), *Oxford Studies in Metaethics* (Vol. 9, pp. 152–185). Oxford University Press.

Cosker-Rowland, R. (2017). The significance of significant fundamental moral disagreement'. *Noûs*, *51*(4), 802–831. https://doi.org/10.1111/nous.12170.

Darwall, S. (1983). *Impartial reason*. Cornell University Press.

Devitt, M. (1983). Dummett's anti-realism. *The Journal of Philosophy*, *80*(2), 73–99. https://doi.org/10.2307/2026236.

Dougherty, T. (2014). Vague value. *Philosophy and Phenomenological Research*, *89*(2), 352–372. https://doi.org/10.1111/phpr.12026.

Dougherty, T. (2018). Vagueness and indeterminacy in ethics. In T. McPherson (Ed.), *The Routledge Handbook of Metaethics* (pp. 185–193). Routledge.

Dreier, J. (1990). Internalism and speaker relativism'. *Ethics*, *101*(1), 6–26. https://doi.org/10.1086/293257.

Dreier, J. (2009). Moral relativism and moral nihilism. In D. Copp (Ed.), *The Oxford handbook of ethical theory* (pp. 240–264). Oxford University Press.

Dummett, M. (1982). *Realism Synthese, 52*(1), 55–112. https://doi.org/10.1007/BF00485255.

Dunaway, B. (2017). Ethical vagueness and practical reasoning. *The Philosophical Quarterly*, *67*(266), 38–60. https://doi.org/10.1093/pq/pqw038.

Dunaway, B. (2018). Realism and objectivity. In T. McPherson (Ed.), *The Routledge handbook of metaethics* (pp. 135–150). Routledge.

Francén Olinder, R. (2016). Some varieties of metaethical relativism. *Philosophy Compass*, *11*(10), 529–540. https://doi.org/10.1111/phc3.12347.

Harman, G. (1975). Moral relativism defended. *Philosophical Review*, *84*(1), 3–22. https://doi.org/10.2307/2184078.

Hawthorne, J. (2005). Vagueness and the mind of God. *Philosophical Studies*, *122*(1), 1–25. https://doi.org/10.1007/s11098-005-0310-4.

Hawthorne, J. (2022). Moral vagueness and epistemicism. *Journal of Ethics & Social Philosophy*, *23*(2), 214–247. https://jesp.index.php/jesp/article/view/1602.

Haybron, D. (2008). *The pursuit of unhappiness*. Oxford University Press.

Johnston, M. (1989). Dispositional theories of value. *Proceedings of the Aristotelian Society Supplementary*, *63*(1), 139–174. https://doi.org/10.1093/aristoteliansupp/63.1.89.

Joyce, R. (2001). *The myth of morality*. Cambridge University Press.

Lewis, D. (1983). New work for a theory of universals. *Australasian Journal of Philosophy*, *61*(4), 343–377. https://doi.org/10.1080/00048408312341131.

Lewis, D. (1989). Dispositional theories of value. *Proceedings of the Aristotelian Society Supplementary*, *63*(1), 113–137. https://doi.org/10.1093/aristoteliansupp/63.1.89.

Mackie, J. L. (1977). *Ethics: Inventing right and wrong*. Penguin.

Manley, D. (2007). Safety, content, apriority, self-knowledge. *The Journal of Philosophy*, *104*(8), 403–423. https://doi.org/10.5840/jphil2007104813.

McDowell, J. (1985). Values and secondary qualities. In T. Honderich (Ed.), *Morality and objectivity* (pp. 110–129). Routledge.

McGrath, S. (2010). Moral realism without convergence. *Philosophical Topics*, *38*(2), 59–90. https://doi.org/10.5840/philtopics201038214.

Miller, A. (2003). *An introduction to contemporary metaethics*. Polity.

Neta, R., & Rohrbaugh, G. (2004). Luminosity and the safety of knowledge'. *Pacific Philosophical Quarterly*, *85*(4), 396–406. https://doi.org/10.1111/j.1468-0114.2004.00207.x.

Olson, J. (2014). *Moral error theory*. Oxford University Press.

Priest, G. (2008). *An introduction to non-classical logic* (2nd ed.). Cambridge University Press.

Putnam, H. (1981). *Reason, truth and history*. Cambridge University Press.

Rachels, J. (1991). Subjectivism. In P. Singer (Ed.), *A companion to ethics* (pp. 423–441). Blackwell.

Ramachandran, M. (2009). Anti-luminosity: Four successful strategies. *Australasian Journal of Philosophy*, *87*(4), 659–673. https://doi.org/10.1080/00048400802587408.

Schoenfield, M. (2016). Moral vagueness is ontic vagueness. *Ethics*, *126*(2), 257–282. https://www.journals.uchicago.edu/doi/pdfplus/10.1086/683541.

Schroeder, M. (2008). *Being for*. Oxford University Press.

Schroeder, M. (2014). The price of supervenience. In his *Explaining the Reasons We Share* (pp. 124–144). Oxford University Press.

Schwitzgebel, E. (2011). *Perplexities of consciousness*. MIT Press.

Sepielli, A. (2012). Normative uncertainty for non-cognitivists. *Philosophical Studies*, *160*(2), 191–207. https://doi.org/10.1007/s11098-011-9713-6.

Shafer-Landau, R. (1994). Ethical disagreement, ethical objectivism and moral indeterminacy. *Philosophy and Phenomenological Research*, *54*(2), 331–344. https://doi.org/10.2307/2108492.

Sider, T. (1995). Sorensen on unknowable obligations. *Utilitas 7*, (2), 273–279. https://doi.org/10.1017/S0953820800002089.

Silk, A. (2018). Metaethical contextualism. In T. McPherson, & D. Plunkett (Eds.), *The Routledge handbook of metaethics* (pp. 102–118). Routledge.

Sobel, D. (2009). Subjectivism and idealization. *Ethics*, *119*(2), 336–352. https://doi.org/10.1086/596459.

Sorensen, R. (1990). Vagueness implies cognitivism. *American Philosophical Quarterly*, *27*(1), 1–14.

Sorensen, R. (1995). Unknowable obligations. *Utilitas*, *7*(2), 247–271. https://doi.org/10.1017/S0953820800002077.

Srinivasan, A. (2015a). Are we luminous? *Philosophy and Phenomenological Research*, *90*(2), 294–319. https://doi.org/10.1111/phpr.12067.

Srinivasan, A. (2015b). Normativity without cartesian privilege. *Philosophical Issues*, *25*(1), 273–299. https://doi.org/10.1111/phis.12059.

Vanrie, W. (2020). Some problems with the anti-luminosity argument. *Pacific Philosophical Quarterly*, *101*(3), 538–559. https://doi.org/10.1111/papq.12317.

Velleman, D. (2015). *Foundations for moral relativism* (2nd ed.). Open Book.

Vogel, J. (2010). Luminosity and indiscriminability. *Philosophical Perspectives*, *24*(1), 548–572. https://doi.org/10.1111/j.1520-8583.2010.00203.x.

Weatherson, B. (2004). Luminous margins. *Australasian Journal of Philosophy*, *82*(3), 373–383. https://doi.org/10.1080/713659874.

Wedgwood, R. (2010a). There are unknowable moral truths. PEA Soup blog. https://peasoup.typepad.com/peasoup/2010/08/there-are-unknowable-moral-truths.html.

Wedgwood, R. (2010b). Moral evil demons. In R. Feldman, & T. Warfield (Eds.), *Disagreement* (pp. 216–246). Oxford University Press.

Williams, B. (1981). Internal and external reasons'. *His Moral Luck* (pp. 101–113). Cambridge University Press.

Williamson, T. (1996). Unreflective realism. *Philosophy and Phenomenological Research*, *56*(4), 905–909. https://doi.org/10.2307/2108292.

Williamson, T. (2000). *Knowledge and its limits*. Oxford University Press.

Wong, D. (1984). *Moral relativity*. University of California Press.

Wong, D. (1991). Relativism. In P. Singer (Ed.), *A companion to ethics* (pp. 442–450). Blackwell.

Wright, C. (1981). Dummett and revisionism. *Philosophical Quarterly*, *31*(122), 47–67. https://doi.org/10.2307/2218677.

Wright, C. (1988). Moral values, projection, and secondary qualities. *Proceedings of the Aristotelian Society Supplementary*, *62*(1), 1–26. https://doi.org/10.1093/aristoteliansupp/62.1.1.

Wright, C. (1992). *Truth and objectivity*. Harvard University Press.

Wright, C. (1996). Response to commentators. *Philosophy and Phenomenological Research*, *56*(4), 911–941. https://doi.org/10.2307/2108293.

Wright, J. (1997). *Realism and explanatory priority*. Kluwer Academic.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.