

ORIGINAL ARTICLE

An argument for moral evidentialism

Marc Andree Weber 

Universität Mannheim, Philosophisches Seminar, Mannheim, Germany

Correspondence

Marc Andree Weber, Universität Mannheim, Philosophisches Seminar, L 9, 5, 68161 Mannheim, Germany.

Email: marc.andree.weber@uni-mannheim.de**Abstract**

Moral evidentialism is the view that one ought *morally* to believe only what is suggested by the evidence at one's disposal. As announced in the title, an argument for (a slightly restricted version of) this view is presented. The argument crucially relies on two specific links between belief and assertion, namely that one should not believe what one must not assert, and that one must not assert what is not suggested by the evidence at one's disposal. In both cases, the normativity involved is moral. The links are established, and it is shown how to deal with seemingly apparent counterexamples.

KEYWORDS

assertion, belief, epistemic normativity, evidentialism, justification, moral epistemology

1 | EVIDENTIAL JUSTIFICATION

Dolores is seriously ill. Her doctor tells her, in slightly smoother words, that she is unlikely to survive the next three months. The doctor also tells her that her chances to recover will improve, at least a fair bit (but not so much as to make survival likely), if she can bring herself to be optimistic and believe that she will defeat her illness.

Should Dolores believe that she will get better? In a way, she should: after all, that increases her chances to survive, and we can safely assume that survival is among her primary aims. This is the *prudential* point of view. In another sense of *should*, however, she should not believe that she will recover, because the evidence she has—her doctor's expert judgement—suggests otherwise. This is the *epistemic* point of view. So it seems that prudential and epistemic considerations pull in different directions regarding the question of what she should believe about her chances to survive.

One could argue, however, that epistemic considerations need not be grounds for Dolores to believe that she will probably die soon. For we aim, epistemically speaking, at acquiring significant knowledge¹ and avoiding to believe falsehoods. Suppose now that Dolores is an able

¹Why 'significant'? Note, for instance, that the first note to this paper contains an odd number of words. Now you know it; but this piece of knowledge is so *insignificant* that you would never aim at acquiring it for epistemic reasons.

scientist who will, if she defeats her illness, continue to do scientific research for many more years. This will probably increase the number of her relevant true beliefs and decrease the number of her false ones. Thus, there is some epistemic merit in her survival, and hence in the belief that supports it. In order to distinguish this epistemic merit from the one that comes from believing what the actual evidence suggests, let us speak of *evidential justification* if what matters is whether someone succeeds at reaching the epistemic aims of acquiring significant knowledge and avoiding to believe falsehoods *right now*, and of *instrumental justification* if what matters is whether someone succeeds at reaching these epistemic aims *in the long run*. Then Dolores is instrumentally but not evidentially justified to believe that she will survive.²

Instrumental justification is concerned with the question of what beliefs will initiate those processes of reasoning that promise to be the most insightful.³ In contrast, evidential justification is solely concerned with the available data and the grounds we have for interpreting them in specific ways and excludes all considerations regarding what beliefs will bring about those processes of reasoning that will best foster our epistemic aims in the long run. So saying that one must believe a proposition only if one has an evidential justification for it means that one is not allowed to believe something against one's evidence, whereas saying that one may believe a proposition on the grounds of an instrumental justification means that one needs to take into account not only one's evidence but also what specific set of beliefs that one adopts now will enable one to reach one's epistemic aims at some given point in time in the future to what degree.

For the purpose of this paper, it is inessential whether instrumental justification is genuinely epistemic. Maybe it is better viewed as a kind of prudential or moral justification with epistemic consequences, because it is not concerned with evidence and its interpretation but with prudential or moral considerations regarding how we will best improve our evidential situation in order to obtain the optimal epistemic consequences for ourselves or others up to a certain point in time. What makes discussing instrumental justification worthwhile is not its putative epistemic character but that we might be permitted in using it to vindicate our beliefs—even if we agree to concentrate on epistemic aims.

The ultimate aim of this paper is to show that from a *moral* point of view, Dolores should believe that she is unlikely to survive the next three months; or, more generally, that from a *moral* point of view, one should base one's beliefs on evidential justification alone (as opposed to, for instance, instrumental justification). This moral obligation might be overridden in some very specific cases; I will point out exactly in which ones.

The argument that proves this is stated in Section 2; its premises are argued for in Sections 3 and 4, respectively. Towards the end of the paper, I will compare my position with the one of the great defender of moral evidentialism, William Clifford (Section 5).

2 | THE ARGUMENT FOR MORAL EVIDENTIALISM

The central principle of evidentialism is that one ought to believe only what is suggested by the evidence at one's disposal. Variants of evidentialism differ, among other things, by the way the obligation is interpreted. Usually, it is taken to be epistemic: one has to believe what is suggested by one's evidence in order to be likely to reach the epistemic aims mentioned above.

²For similar cases, see Conee (1987) and Matheson (2014, 2015, pp. 145–146). Matheson also draws a distinction between evidential and instrumental justification—only that he uses the terms 'synchronic (epistemic) justification' and 'diachronic (epistemic) justification' to signal that diachronic justification is a kind of epistemic justification. As I prefer not to assume this (see below), I choose a terminology that is more neutral with respect to the question of whether instrumental justification is epistemic or prudential or both.

³Of course, the processes of reasoning that will prove most insightful within the next, say, 5 years might be others than those that will prove most insightful within the next, say, 50 years. We can leave it open here at which future point in time we will take stock.

A minor problem here is the inexactness of the common use of ‘epistemic’ that I described in section 1: whereas it is usually understood that epistemic aims are to be reached *right now*, it is rarely explicitly ruled out by epistemic evidentialists that instrumentally justified beliefs are a subset of epistemically justified beliefs. However, if instrumentally justified beliefs form such a subset, then there can be epistemic norms—namely, norms to form instrumentally justified beliefs—that contradict the evidentialist’s principle to believe what is suggested by one’s evidence. This would give epistemic evidentialists (those who hold that one *epistemically* ought to believe only what is suggested by the evidence at one’s disposal) a hard time, for then they would have to find arguments for undermining instrumental justification in order to defend their view. If, on the other hand, instrumentally justified beliefs belong to the prudentially justified ones, they do not concern *epistemic* evidentialism; they rather add to the more general problem of balancing epistemic and prudential norms. As my concern here is not with epistemic but with moral normativity, I need not decide whether or not to subsume instrumental under epistemic justification.

A more pressing problem for epistemic evidentialists concerns the reasons we have for pursuing epistemic aims at all. Is it an end in itself to do so, perhaps because it is implied by the very concept of belief that beliefs aim at truth?⁴ Or do we need another kind of normativity to ‘back up’ epistemic normativity? If the latter, we would presumably need either prudential or moral grounds for pursuing epistemic aims, and the central evidentialist principle would then be justified by recourse to either our own well-being or to moral rightness. In the first case, we get prudential evidentialism, in the second case moral evidentialism.⁵

Moral evidentialism is thus the thesis that one ought morally to believe only what is suggested by the evidence at one’s disposal, or, in other words, only what one is evidentially justified to believe. What is meant by this is not that evidential justification is a necessary condition for the moral obligation for belief, but that it is a moral obligation to take evidential justification as necessary for belief; the ‘ought’ here refers to the whole conditional, not only to the antecedent. More perspicuously, we could state the thesis of moral evidentialism, of which I will vindicate a restricted version, as follows:⁶

ME One must (morally) : believe $[p]$ only if one is evidentially justified to believe $[p]$.

Note that this is not a biconditional: ME does not require one to believe $[p]$ if one is evidentially justified to believe $[p]$. A consequence of this is that Dolores is not forced by ME to believe that she will probably die soon, although her evidence strongly suggests this. However, this turns out to be a poor consolation for her, because if she considers the proposition whether she will survive her illness at all, she cannot take any other stance to it than the one suggested by her evidence. The reason is that she is neither allowed to believe that she will recover (for her evidence clearly suggests that she will not survive), nor is she allowed to believe that it is hard to

⁴Many epistemologists hold that something like this is true. Some argue that beliefs are intrinsically normative, so that a mental attitude that comes without normative force cannot be a belief, and that doxastic norms are conceptual truths (Adler, 2002). Others seek to establish a necessary connection between epistemic judgements and epistemic motivation (Mitova, 2011; see also Boulton & Köhler, 2020 for discussion). Still others suggest that it is a psychological phenomenon that we, whenever we consider whether we should believe some proposition, consider what speaks for or against this proposition’s truth, and that the best explanation for this phenomenon implies that only evidential reasons are admissible reasons for belief (Shah, 2006). Still others claim that reasons for belief must be ‘premises of good reasoning’ (Way, 2016); as a consequence, someone who believes a proposition to be true for non-evidential reasons commits a performance error. All of these approaches locate the source of normativity either in the concept of belief or in the way our belief-forming faculties, such as perception or reasoning, are set up. It is a matter of ongoing discussion whether any of these accounts succeeds in establishing the alleged fundamentality of epistemic normativity.

⁵For an argument in favour of prudential evidentialism, see Reisner (2009). Moral evidentialism is famously advocated in Clifford (1877/1999); contemporary vindications include Wood (2008) and Ryan (2015). Both Wood and Ryan defend moral evidentialism against various alleged counterexamples; Wood also argues that self-respect and respect for others require us to subscribe to it. As their lines of reasoning are obviously very dissimilar to mine, I will refrain from any comparison.

⁶Throughout this paper, ‘ $[p]$ ’ is shorthand for ‘the proposition that p ’.

tell whether she will recover (for her evidence clearly suggests that she will not survive), nor is there any other relevant belief, apart from the one that she will probably die soon, that is compatible with her evidence.⁷ Dolores can avoid the belief that she will probably die soon only by violating ME or by avoiding to think about her prospects for survival at all—hardly a feasible option in her situation. The general lesson is that even if ME does not prescribe that one should believe what one's evidence suggests, it excludes all relevant doxastic alternatives, so that the only way to both comply with ME and avoid believing what one's evidence suggests is to believe nothing at all. (Of course one could argue that once one is aware of the relevant evidence, one could not abstain from forming any belief at all, for conceptual or psychological or perhaps even moral reasons. If for moral reasons, this would establish the biconditional. In this paper, however, I will be content with ME, for this sufficiently restricts what we are allowed to believe.)

Here, then, is (a preliminary version of) the main argument:

- (1) One must (morally): believe $[p]$ only if one is (morally) allowed to assert $[p]$.
 - (2) One must (morally): assert $[p]$ only if one is evidentially justified to believe $[p]$.
- ∴ ME Therefore, one must (morally): believe $[p]$ only if one is evidentially justified $[p]$ to believe $[p]$.

The first premise states a norm of belief; the second premise states a norm of assertion. In formulating the norms of belief and assertion, an 'only if' construction was chosen in order to expose more clearly that necessary conditions for belief and assertion are stated.

As it stands, the argument is not sound because both premises are wrong. Two things are missing. First, in each case, we need a normality constraint that filters out cases in which certain other obligations get in our way. As these other obligations are epistemically irrelevant, we can derive the conclusion nevertheless. Second, we need to restrict the second premise and a fortiori the conclusion to cases in which the relevant belief is not directly causally efficacious to a high degree. As cases of such direct causal efficaciousness are rare, the restriction is not severe. I will point out below (in Sections 3.2, 3.3, and 4.2) how the premises and the argument need to be modified. For simplicity's sake, I state here the more lucid unrestricted version.

Conditional norms such as (1), (2), and ME, in which it is not up to the agent whether the relevant condition holds, are often taken to have narrow scope; they are usually formalised along the lines of what some authors call the 'bridge conception' (see e.g. Navarro & Rodríguez, 2014, pp. 92 and 100). This means that only the consequent appears within the scope of the deontic operator; the correct logical form is $p \rightarrow \mathbf{O}\phi$ (where p is a proposition, ϕ an action or, more broadly, anything that can be done for a reason, including believing something, and \mathbf{O} the deontic operator for obligation). As one cannot in the least control what one is allowed to assert, or evidentially justified to believe, since because this depends on norms that are independent of one's will, it is reasonable to apply the bridge conception in analysing the conditional norms (1), (2), and ME.

By contraposition, ME is equivalent to *one must (morally): not believe $[p]$ if one is not evidentially justified to believe $[p]$* . In accordance with the bridge conception and with the obvious abbreviations, this proposition can be formalised as $\neg J_e Bp \rightarrow \mathbf{O}\neg Bp$. By defining permissibility in the usual way, i.e., $\mathbf{P}\phi \leftrightarrow_{\text{def}} \neg \mathbf{O}\neg \phi$, and by using contraposition again, $\neg J_e Bp \rightarrow \mathbf{O}\neg Bp$ is equivalent to $\mathbf{P}Bp \rightarrow J_e Bp$. In the same way, (1) is equivalent to *one must (morally): not believe $[p]$ if it is not allowed to assert $[p]$* , which can be formalised as $\neg \mathbf{P}Ap \rightarrow \mathbf{O}\neg Bp$ or,

⁷This argument presupposes rational uniqueness. Rational uniqueness is defended, e.g., in White (2005), Christensen (2007), and Feldman (2007) (although only Feldman explicitly acknowledges what he calls the 'Uniqueness Thesis'). Although not logically entailed by my definition of evidentialism, uniqueness is naturally seen as an integral part of it.

equivalently, as $\mathbf{PB}p \rightarrow \mathbf{PA}p$; and (2) is equivalent to *one must (morally): not assert $[p]$ if it is not evidentially justified to believe $[p]$* , which can be formalised as $\neg J_e Bp \rightarrow \mathbf{O}\neg Ap$ or, equivalently, as $\mathbf{PA}p \rightarrow J_e Bp$. In sum, we can formalise the argument as follows:

$$\begin{aligned} & \mathbf{PB}p \rightarrow \mathbf{PA}p \\ & \mathbf{PA}p \rightarrow J_e Bp \\ \therefore & \mathbf{PB}p \rightarrow J_e Bp \end{aligned}$$

By the transitivity of the conditional, the argument is valid. Before I shall argue, in Sections 3 and 4, that its premises are true (or almost true), three remarks are in order.

First, premise (1) links assertion and belief in an unusual way insofar as it uses the former to restrict the latter. This is astonishing because belief, unlike assertion, is commonly thought of as something private, which involves duties the believer has to oneself but not duties the believer has to others. Whatever responsibilities to audiences and maxims of conversation govern assertions, they seem to have little to do with the norms for belief. Thus, at first glance, this premise appears rather implausible.

Second, the argument is peculiar in that its conclusion may deem many philosophers plainly more plausible than its first premise (and maybe even a fair bit more plausible than its second one). This might partly be due to the fact that an epistemic variant of ME seems hard to deny, partly due to the close connection between belief and evidential justification, which lets even the moral variant appear quite defensible. Yet ME, and even an epistemic variant of it, might be reasonably doubted in the light of the prudential and instrumental reasons that Dolores has for the belief that she will survive, and at least the moral variant should be taken into question as long as we have not seen any positive reason for it, because moral considerations are normally taken to govern actions or intentions, but are not standardly extended to cover beliefs as well. Yet they should cover beliefs, and the premises explain why.

Third, the argument may appear redundant, at least implicitly. The reason is that it first links belief with assertion, and then assertion with belief, so that it seems as if we use the same connection twice. The idea here is that what might appear questionable on the level of beliefs, namely that they ought to be evidentially justified, appears less questionable on the level of assertions. So by establishing a norm of belief that links the permissibility of beliefs with the permissibility of assertions, namely (1), we are able to use a not-too-controversial fact about assertion, namely (2), to derive a non-standard (although perhaps initially plausible) conclusion that concerns beliefs.

3 | PREMISE (2): A NORM OF ASSERTION

3.1 | Rationale for (2)

Let us start with the second premise—*one ought (morally) to assert $[p]$ only if one is evidentially justified to believe $[p]$* —and let us ignore for a moment the nature of the normativity. The claim that one must assert $[p]$ only if one is evidentially justified to believe $[p]$ presupposes two non-trivial theses about assertion, namely that the speech act of assertion is governed by a norm, and that this norm entails evidential justification. Both assumptions are not uncontroversial, but widely held.

The idea that assertions are governed by a norm is the idea that one can give a rule, or a set of rules, such that exactly those speech acts are governed by this rule, or set of rules, that are assertions. This is not to say that all assertions are in accordance with these rules; rules can be

broken, and often are. But it means that there is something faulty in making an assertion that violates the rules. Arguably, this explains the several features of assertions, including the ways in which an assertion might be misplaced, better than rival accounts, such as the view that assertions are mere expressions of beliefs and/or intentions that the hearer forms the respective beliefs, or the view that in asserting a proposition, one undertakes a commitment to its truth.⁸

Assume, following Williamson (2000, p. 241), that there is exactly one constitutive rule of assertion, which has the form ‘one must: assert $[p]$ only if $[p]$ has C .’ Candidates for $[p]$ having C include that

- (i) $[p]$ is true;
- (ii) one has warrant to assert $[p]$;
- (iii) one knows $[p]$;
- (iv) one justifiably believes $[p]$;
- (v) one rationally believes that one knows $[p]$;
- (vi) one is in a position to know $[p]$.

Of these candidates, (iii), (iv), and (v) involve that one believes $[p]$, and all but (i) involve that one has some sort of justification for believing $[p]$. As a unique rule, however, (i) is implausible because it appears blameworthy to assert a proposition without having any evidence for its truth, even if this proposition coincidentally turns out to be true.⁹

Far from being conclusive, these sketchy remarks may at least indicate that whatever norm governs assertion, it must imply that having some sort of justification for believing $[p]$ is a necessary condition for asserting $[p]$. What I claim in addition to this is that this justification needs to be *evidential*. This is not trivial; for in (ii)–(vi), one could argue that the justificatory element might partly be understood in terms of instrumental justification. Even in the case of knowledge, one could claim that it is sufficient for an ungettiered true belief to be instrumentally justified in order to be regarded as a piece of knowledge.

So what we have to show is that an evidential but not an instrumental justification for a belief is necessary for the permissibility of the respective assertion. To prepare the ground for the proof, suppose that another person is in need of information regarding whether p , and asks you to tell her whether p is true. Suppose further that there are, as far as you know, no reasons for you not to help her: you do neither dislike her, nor disbelieve that she really needs to know whether p , nor are you short of time, etc. Given that it is epistemically permissible for you to assert either $[p]$ or $[\neg p]$ —that is, you have sufficient epistemic warrant either for the truth of $[p]$ or for its falseness to make the respective assertion—you should assert either $[p]$ or $[\neg p]$ and thereby meet the need for information of the person who asked you. The obligation here is moral, not epistemic: another one is in need of help, you are in a position to help, helping comes without any costs to speak of, hence you should help. (If you are unconvinced, imagine the required help to be of utmost importance, your certainty that you can provide it virtually absolute, and the effort for you utterly negligible.)

In short, assertion involves a moral dimension that obliges the speaker to address the hearer’s informational needs as long as they are not overridden by other considerations. This is how moral normativity comes in. By contrast, the normativity in (i)–(vi) is generally not considered moral. It is better taken to be the same kind of constitutive normativity that governs our behaviour in games such as football (see Williamson, 2000, p. 240). The rules of football regulate what players are allowed to do on field. If a rule is broken—which frequently happens—the offender is penalised or, if not, would have been penalised if the referee had noticed the rule

⁸See Goldberg (2015, pp. 9–12), drawing on MacFarlane (2010), for an overview of the main accounts of assertion, and Goldberg (2015, pp. 6–35), and Williamson (2000, pp. 238–243), for defenses of a norm-based account.

⁹For a more extensive argument, see Williamson (2000, pp. 244–249).

violation. By analogy, the rule of assertion regulates what speakers are allowed to assert to each other, and a speaker who violates this rule thereby does something blameworthy, so that those who notice the violation could rightly criticise the asserter. To be sure, violating constitutive rules may often have some moral dimension (cheating, for example, is usually considered morally wrong), but this is not normally taken to be central for the normative character of those rules.

Let us go back to the claim that, other things being equal, one should address another person's informational needs. The essential point to complete our proof is that one cannot in general do this adequately by telling the other person what one is only instrumentally justified to believe. Recall, for illustration, Dolores, the deathly ill scientist. If she is asked by a friend about her survival prospects and tells him that she is likely to recover, she thereby violates her moral duty to give him the requested information.¹⁰ She is not necessarily lying, because she may have succeeded in acquiring not the evidentially but the instrumentally justified belief, and thus just tells her friend what she actually takes to be the case; but her failure is that she has either allowed herself to slip into a state of mind in which she is incapable of answering other people's questions regarding her health properly, or that, while still aware what answer would be appropriate, she rather asserts what she believes (perhaps in order to prevent her social environment from questioning her belief). In both cases, she wrongly creates the impression of satisfying informational needs.¹¹ (Note that my point here is merely that we can dismiss for purely moral reasons an interpretation of the evidential element in (ii)–(vi) that involves instrumental justification. This does not mean that there are no other reasons to dismiss such an interpretation—one might, for all I see, convincingly argue for a constitutive norm of assertion that explicitly excludes instrumental justification. It does only mean that, perhaps in addition to such a constitutive norm, there is a distinctly moral norm.)

3.2 | Morally permissible lies

There is an obvious objection. Assume that I would cause much harm by not asserting something that I do not justifiably believe. For instance, assume that I would betray the whereabouts of innocent people to the henchmen of a cruel dictator if I did not lie to them. This appears to be a case in which I am morally allowed—perhaps even required—to assert a proposition (here, that those people are at place *X*) that my evidence clearly does not support. Hence, we are forced to revise (2) to

- (2*) If there are no overriding moral considerations, one must (morally) : assert [*p*]
only if one is evidentially justified to believe [*p*].

It is easy to see, however, that this revision is harmless for my overall argument for moral evidentialism. For (2*) is only a reasonable alternative to (2) if we refer with 'overriding moral considerations' to moral factors that influence what one is allowed to assert but clearly not what one is evidentially justified to believe. More precisely, moral norms of assertion that are to

¹⁰She would not, of course, violate this duty by saying something like 'Well, the doctor says I won't make it. But I *will* make it' but only by saying something like 'I'm likely to make it' (and not adding anything). If she makes her evidence transparent, as in the first case, she enables her audience to draw their own conclusions, so that they do not rely on her belief statement any more. Note, however, that if she trusts her belief, or wants others to trust it, she has no reason to disclose her evidence and conceal what the doctor told her. (I am grateful to an anonymous reviewer for pressing me on this point.)

¹¹If there are prudential reasons for Dolores not to address her friend's informational needs, e.g., because he cannot be trusted, or would annoy her by pitying her excessively, she may not make any relevant assertion, but is not allowed to assert something that she has no evidential justification for.

be filtered out by the ‘overriding moral considerations’ clause must not give us an evidential justification to believe differently than we otherwise would. Norms governing lies are a case in point, for they may have an impact on what we are allowed to say to other people, but do not affect how we should form our beliefs. Hence, whatever overriding moral considerations there might be, they have nothing to do with evidential justification.¹²

3.3 | Directly causally efficacious beliefs

There is, however, a variant of the case of morally permissible lies that causes real trouble.¹³ Assume there is an artist whose work I find (after evaluating it carefully) very poor, but who threatens to kill dozens of innocent people if I do not consider it brilliant. Unfortunately, the artist has a fail-safe mind-reading device, so that he will know exactly what I believe. In order to save those innocent people from death, it does not suffice that I lie about my beliefs concerning the quality of his art; I also have to bring myself to believe that his art is great. Hence, this appears to be a case in which I am not only morally permitted—perhaps even required—to assert what is not supported by my evidence (i.e., my visual impressions of his art and my infallible sense of aesthetic beauty), but in which I am also morally permitted or required to believe against my evidence.¹⁴

The crucial feature of this case is the fail-safe mind-reading device. This feature makes my belief causally efficacious in a direct way, whereas usually, beliefs have consequences only indirectly, by helping a person to arrive at a decision of what to say or do. Due to this direct way of causation, the only way to lie to the artist is by deceiving myself. Such self-deceptions, albeit possible at least in principle (see Section 4.3), are normally morally doubtful due to the danger of deceiving others as well. In the present case, however, the moral aim to deceive the artist requires me to deceive myself.

We can imagine further variants in which the artist has no fail-safe way of reading my thoughts but is excellent at detecting lies, or in which I am a very poor liar. In such cases, my beliefs are less directly causally efficacious than in the original case, so that attempts to disguise them might turn out successful. Nevertheless, it may still be safer not to have the beliefs in question. In the light of the severe consequences, I might therefore be obliged to avoid them. Realistically, however, in these variants, the artist will not be sure about whether he is told lies—and this holds both when he is in fact lied to and when I deceived myself and am now sincere. Hence, what the artist will do depends much more on his proclivities in cases of uncertainty as on my actual belief. This decreases considerably the moral importance of self-deception in cases in which my beliefs are less directly causally efficacious.

My beliefs may also be directly causally efficacious in other ways, for example, by making me do something, by giving me motivation to persevere, or by making me biased in some way.¹⁵ If there are strong moral reasons for or against doing the respective actions or having the respective motivations or biases, self-deception might be morally required. However, I think we tend to overestimate the causal power of mere beliefs here: if there are indeed strong moral reasons for acting in a certain way, we normally have enough strength of will to do our very best and overcome whatever biases we know ourselves to have (and by self-deception, we could only guard us against such biases, not against those that have been unknown to us). It is usually not the beliefs that are to blame, but weakness of the will. And when the alternative of staying strong is open to us as well, we should not deceive ourselves.

¹²Even so, the substitution of (2) with (2*) makes it necessary to add an auxiliary premise. See section 4.2 for details.

¹³This case is presented and discussed in Ryan (2015). See also Brogaard (2014, p. 132).

¹⁴Note that I am not epistemically permitted to believe against my evidence; for there is no epistemic reason to save the innocent people. The scenario does not work as a counterexample to epistemic evidentialism.

¹⁵See, e.g., Meiland (1980) and Heil (1983) for classical cases in this vein.

As an illustration, consider a variant of the case of deathly ill Dolores. In the original case, Dolores has a prudential and perhaps an instrumental justification for believing against the evidence, but not a moral one. Now let us assume that some mad person convincingly threatens to kill dozens of innocent people if Dolores dies from her illness. Then Dolores is also morally obliged to do her best to survive. Does this include that she should form the respective belief? I do not think so. The reason can be seen by a more realistic understanding of the scenario. For why does the belief that she will defeat her illness foster Dolores's survival? Presumably because it gives her the strength to do everything she could to survive: endure stressful therapies, eat well, lead an everyday life as normal as possible, keep her job, make plans for the future, and so on. It is not the very belief that renders her survival more likely, but its probable effects, the way it motivates Dolores to do what might help her. Hence, Dolores is morally required to endure stressful therapies, eat well, and so on, but not to believe that she will get better soon. She would only be morally required to form this belief if otherwise she were not able to bring herself to do all these things due to weakness of will. This is why beliefs like this one are usually not directly causally efficacious, or at least not to a high degree.

The upshot of these considerations is that although direct causal efficaciousness appears to come in degrees, self-deception may only be morally permitted in clear cases of direct causal efficaciousness. But these cases indeed force us to restrict (2*) as follows:

- (2***) If believing $[p]$ is not directly causally efficacious to a high degree and if there are no overriding moral considerations, one must (morally) : assert $[p]$ only if one is evidentially justified to believe $[p]$.

Almost always one's beliefs are causally efficacious in the indirect way, or largely so. Hence, this restriction is not devastating for applying moral evidentialism. Yet it shows that ME does not hold fully generally, as the restriction stated in the first if-clause in (2***) carries over to ME (see Section 4.2).

3.4 | Directly morally efficacious beliefs

One's beliefs are not only causally but also *morally* efficacious in an indirect way: under suitable circumstances, they give us reasons to act in certain morally evaluable ways. For example, if I believe that my stopping to eat meat helps to significantly reduce animal pain, this gives me a reason to become a vegetarian.

According to what I am arguing, one's beliefs are also directly morally efficacious: the mere formation or maintenance of evidentially unjustified beliefs already is morally wrong since because it excludes the option of giving others appropriate information. Some philosophers have claimed that there are further ways in which one's beliefs are directly morally efficacious; that even if they are evidentially justified, merely forming or maintaining them may wrong other people. To give but four examples:

- a. Basu and Schroeder think that certain negative albeit evidentially justified beliefs harm other people by discriminating or downgrading them (Basu, 2019b; Basu & Schroeder, 2019). If I believe, for instance, of a Black man in a noble restaurant that he is a waiter rather than a guest, since because this is, as the example goes, the much more likely role for a Black person in this restaurant, then my very belief wrongs that person.
- b. Stroud argues that being a good friend can require epistemic partiality: part of what it means to be a good friend is to resist thinking badly about a friend even if the evidence suggests

- otherwise (Stroud, 2006; see also Keller, 2004). You should, for example, disbelieve some worrying story about a friend even if there is ample reason to trust that story.
- c. Marušić claims that the social practice of making sincere promises requires the promise-giver to believe that she will keep her promise even if this is unlikely (Marušić, 2013, 2015). If I promise, for example, that I will stop smoking, reliable statistical evidence speaks against my fulfilling this promise; nevertheless, I can only sincerely promise to quit smoking if I believe, contrary to the evidence, that I will succeed.
 - d. Alfano suggests that considering others to be virtuous may help make them so and may therefore be called for even in the absence of evidence (Alfano, 2013, chap. 4). If you think, for example, that I am trustworthy and treat me accordingly although your evidence does not sanction your belief, this may motivate me to live up to your belief.

While Basu and Schroeder emphasise that evidentially justified beliefs that are discriminating or downgrading should not be had for *moral* reasons, it is less clear in the other cases which kind of normativity should be taken to speak against having evidentially justified beliefs. One option is to maintain that epistemic evidentialism is correct, but to point out that considerations on friendship, promises, or self-affecting virtue ascriptions show that social norms require us to form evidentially unjustified beliefs. Whether social norms trump epistemic ones or cannot be balanced against them is then a matter of further discussion. Another option is to claim that (b)–(d) give us epistemic reasons to avoid certain evidentially justified beliefs—just as instrumentally justified beliefs can be said to give us such epistemic reasons—and to conclude that epistemic evidentialism is false. A still other option, which is neither held by Stroud, Marušić, or Alfano but could be vindicated with some plausibility, is that (b)–(d) indicate that it is morally wrong to form certain evidentially justified beliefs, either because social norms that govern friendship, promises, and virtue ascriptions are grounded by moral norms, or because of some specific, perhaps consequentialist argument in favour of an immediate moral demand for epistemic partiality in (b)–(d). Only according to this last option, (b)–(d) seem to contradict moral evidentialism (and more specifically, my premise (2**)) in the same way as (a) does.¹⁶

There are, as far as I see, three lines of defence available for moral evidentialism. First, one could advocate *moral encroachment*, the view that what is evidentially justified depends at least partly on what is morally at stake,¹⁷ and then claim that the risks of discriminating or downgrading someone, of not being a good friend, of being unable to sincerely make substantial promises, or of missing chances to improve other people's virtues affect the moral stakes. As a consequence, the descriptions of the examples in (a)–(d) would be misleading insofar as the protagonists' evidence in those cases does actually *not* suffice to make them evidentially justified (or to work as a defeater for another belief, for instance one's belief that one will keep a certain promise); additional evidence would be needed for evidential justification (or for defeat).

Second, one could argue that the moral significance of addressing other people's informational needs is far greater than whatever moral significance there is for having evidentially unjustified (and not directly causally efficacious) beliefs. Part of this argument would be an evaluation of the respective moral significances of such examples as given in (a)–(d).

The third way to vindicate moral evidentialism is by arguing that in each of the cases (a)–(d), there are case-specific reasons why it is false to think that one is morally required, in that case, to believe against one's evidence. For instance, one could argue that our qualms to draw

¹⁶Remember that (2**) is a norm of assertion according to which evidential justification is *ceteris paribus* necessary for assertion, where the *ceteris paribus* clause filters out cases of directly causally efficacious beliefs and counterexamples that are irrelevant for what we are allowed to believe. Now assume that I, given that I promise to you to stop smoking, am *morally* allowed to believe and also to assert, contrary to my evidence, that I will stop smoking. Then the moral permissibility of this assertion would violate (2**), as the corresponding belief is neither directly causally efficacious nor evidentially justified.

¹⁷For arguments in favour of moral encroachment, see, e.g., Pace (2011), Basu and Schroeder (2019), and Basu (2019a). For discussion, see, e.g., Fritz (2017), Gardiner (2018), and Hirvelä (2023).

negative conclusions about a friend are better explained by the detailed evidence one presumably has of one's friend's good character; ignoring one's positive valuation of the friend's character would then amount to disregarding relevant evidence. As a result, being a good friend would be compatible with evidentialism (in either form).¹⁸

I cannot do justice to either line of defence here; this would require several additional papers. My aim in this paper is merely to present an argument in favour of moral evidentialism, not to refute every argument against it. So, I rest content with concluding that while (a)–(d) might appear to present counterexamples to the restricted version of moral evidentialism that I am endorsing, there are several theoretical options to deal with these alleged counterexamples.

3.5 | Long-term epistemic aims

There is another objection: what is meant by 'informational needs' is not yet sufficiently clear. In our original scenario, we said that Dolores's survival will enable her to continue doing scientific research so that *her* epistemic aims will probably be fulfilled to a greater degree in the long run. Assume, in addition, that her future research might easily turn out important (perhaps she is a leading expert in her field) so that the epistemic aims of *many others* will probably be fulfilled to a greater degree in the long run if she survives. And among the many others who profit might be her friend who asked her about her survival prospects. So, arguably, telling him that she will survive might serve his long-term epistemic aims and can thus be considered to meet his informational needs. In that case, Dolores would be entitled to assert what she had only an instrumental justification for. Put differently, the expression 'epistemic warrant' in 'having sufficient epistemic warrant either for the truth of [*p*] or for its falseness to make the respective assertion' should be understood in terms of instrumental justification if the informational need is better addressed by telling what serves long-term epistemic aims than by telling what serves short-term ones; and the Dolores case indicates, according to the objection, that it might well happen that informational needs are indeed better addressed by telling what serves long-term epistemic aims.

Yet the objection is misleading, as it becomes clear when we consider cases in which the informational needs have broader or more radical consequences than the informational needs of Dolores's friend, or in which they affect institutionalised customs. For instance, it would have appeared insincere to us if Dolores's doctor had told her, in order to support her long-term epistemic aims, that she will probably recover. The reason is that we take it for granted that this information is based on evidence, not on strategic aims, and that a violation of this expectation is the less excusable the stronger the influence is that it has on the future actions of the addressee. Even more blatant is the dishonesty in cases in which someone who makes an assertion claims to present results of scientific research. Consider the following case:

Global Warming is a serious threat. It would be great if we knew as much as possible about its mechanisms, the effectiveness of counter-measures, potential climate tipping points, and so on. Now, the most promising way to acquire valuable new knowledge about such matters is to raise a lot of research funding; and the most promising way to do that is to persuade people and politicians that the risks of global warming are even much greater than they in fact are (to the best of our current knowledge). Therefore, climatologists are instrumentally justified to assert propositions that contradict their evidence, for that would help them to fund their future research, and would thereby contribute to acquire more and better knowledge in the long run.

¹⁸Stroud mentions this option as well but thinks that recourse to character evidence does not suffice to explain the relevant examples. See Stroud (2006, pp. 515–516).

The instrumental justification of these climatologists clearly does not entitle them to assert something that is not in line with their evidence. The more we expect people to base their assertions on careful analyses of the evidence, as scientist ought to do, the less we find it tolerable if they betray our anticipation that what they say reflects their evidence.

The cases in which it is most obvious that instrumental justification should not affect what we assert are those in which the assertion comes with a certain emphasis, for instance, because much depends on its content or because it is made with scientific authority. This is telling, for these are exactly the cases in which we have much reason to care about whether the rules that govern assertions are observed. If we feel with regard to these cases that such rules are violated by those who assert a proposition on the basis of the instrumental justification they have for it, then those rules seem to forbid assertions based only on instrumental justification *tout court*. Therefore, we should abstain from interpreting the evidential element in (ii)–(vi) in Section 3.1 in terms of instrumental justification.

4 | PREMISE (1): A NORM OF BELIEF

4.1 | Rationale for (1)

This premise—*one ought (morally) to believe*[*p*] *only if one is (morally) allowed to assert* [*p*]—implicates that there is, in addition to a moral norm of assertion, a moral norm of belief, and that the latter is at least as strong as the former. So, necessarily every scenario in which one is allowed to believe something but not allowed to assert it violates this norm of belief. The opposite, i.e., that the assertion but not the respective belief is permissible, is not ruled out.

Several arguments can be found in the literature for a version of (1) that involves a non-moral ‘ought.’¹⁹ For different reasons, which lack of space forbids me to elucidate here, I find these arguments unconvincing; a moral version of (1), although perhaps initially less plausible, seems to me better justified. Like a non-moral version, however, a moral version is exposed to a variety of seemingly obvious counterexamples. In what follows, I will first state an argument for (1) and then point out how we can deal with the counterexamples.

Assume that one believes [*p*]. I shall argue that if the belief is permissible, this already gives one sufficient warrant to assert [*p*]. Observe first that to believe [*p*] is to believe that [*p*] is true; for if one comes to think that what one believes is false, one thereby stops believing it (see Williams, 1973, p. 137). One could add that if one comes to think that one has no reasons for believing [*p*], one thereby stops believing it.²⁰ So, if one believes [*p*], one necessarily has reasons (or believes to have them) to think that [*p*] is true, or, put differently, one has to have, for purely conceptual reasons, some kind of truth-aiming justification for [*p*].²¹ Note that this justification need not be a good one: if the belief has resulted from a careless or superficial evaluation of one’s evidence, and if one is not aware of this lack of thoroughness, then one takes oneself to have a good justification without, in fact, having one.

Second, the most natural kind of a good justification (and perhaps the only one) is evidential justification. Being evidentially justified to believe [*p*], however, normally sanctions asserting

¹⁹See Bird (2007, p. 95), Sutton (2007, p. 46), Kvanvig (2009), and Maitra and Weatherston (2010, pp. 104–105). For an overview, see Goldberg (2015, pp. 163–164).

²⁰The notion of a *reason* is to be understood in a very broad sense here, according to which things like experiences, anticipations, intuitions, premonitions, gut feelings, informed guesses, and so on can give us reasons as well. All that is needed is some trust in one’s respective abilities.

²¹Arguably, I might believe [*p*] mindlessly, without being aware of any supporting reasons. While I do not think that this is possible—how can I form a belief if there is not even some vague feeling that speaks in favour of it?—it would not be problematic for my argument for premise (1) because if it is morally wrong to believe anything for the wrong reasons it certainly is morally wrong as well to believe anything for no reasons at all (Thanks to an anonymous reviewer for pressing me on this point.).

[*p*] because of the obvious epistemic benefit other people generally have if their explicit or implicit informational needs are addressed by the assertion of a proposition for which there is a non-defeated evidential justification. So, if we will disprove (1), we must assume that one might have a kind of truth-aiming justification for [*p*] different from, and not including, evidential justification. (Keep in mind that this justification need not be a good one, as long as one takes it to be good.)

Third, a non-evidential justification necessarily is either truth-disguising rather than truth-tracking, or it is not.²² If it is truth-disguising, one should abandon one's belief in favour of a belief backed up by a justification that is in some form or another truth-tracking; otherwise this would be a case of self-deception (see Section 4.3). If it is not truth-disguising, it indeed entitles one to believe [*p*], but does so in a way that is independent of one's specific point of view, because if truth is not relative to points of view (a standard assumption, which I will not discuss further here), truth-tracking justification is not relative to points of view either, and one's truth-tracking justification for [*p*] would then be a truth-tracking justification for [*p*] for other people as well. This, again, sanctions asserting [*p*] because of the obvious epistemic benefit other people have if their informational needs are addressed by the assertion of a proposition for which there is a non-defeated, truth-tracking, truth-aiming justification (or seems to be, if one is unaware of one's lack of thoroughness). Therefore, believing [*p*] suffices to entitle one to assert [*p*], except in cases of self-deception; in these cases, people form or maintain beliefs that they are not morally permitted to form or maintain, as I will argue below. In short: due to their conceptual tie to truth, the beliefs that people actually have are automatically justified (or at least appear so to the believers) as long as no self-deception is involved. The justification here is either evidential or, if there can be such a thing, non-evidential but truth-tracking. What is more, it is morally permissible to form or maintain beliefs that are justified in this way, as they do not normally cause any epistemic harm to other people.

(Of course, there are all kinds of sexist or racist beliefs, delusions, cases of wishful thinking and so on; but the people who have those beliefs usually are, from their own points of view, justified to have them because they think those beliefs track the truth. For example, a racist does not only believe that some races are superior to others but also that one is justified in taking some races to be superior to others. The belief in this justification either results from having been exposed to misleading evidence, or from not having evaluated one's evidence with sufficient care, or from having successfully deceived oneself. In the first case, one is justified (in the internalist sense employed here throughout) to assert one's beliefs. In the second case, one has no justification for one's beliefs and should abandon them. For the third case, see Section 4.3.)

Finally, the entitlement for assertion that one gets from an evidential or a non-evidential truth-tracking justification is moral. The reason is that asserting a proposition in which one has a non-misleadingly justified belief does not normally cause any epistemic harm to other people (for instance, via misrepresentation of facts or their misinterpretation), and not causing epistemic harm by wrongly addressing other people's explicit or implicit informational needs is the crucial moral norm that must be obeyed in connection with assertion. Therefore, asserting what one justifiably believes is morally permissible.

There seem to be several counterexamples to (1): scenarios in which justified belief appears to meet disallowed assertion. They involve conflicting norms, self-deception, and retraction, respectively. The replies to these counterexamples are essential for a complete understanding of the argument. I will consider them in turn.

²²There is a difference between truth-aiming and truth-tracking justifications. For instance, if I believe in the prophecies of a fortune teller, the justification for my beliefs—'the fortune teller has said so!'—is based on truth being the ultimate aim of believing, and thus is truth-*aiming*, but in fact directs me away from forming true beliefs, and thus is not truth-*tracking* (and I could have known it not to be so).

4.2 | Conflicting norms

This is the most obvious kind of counterexample, which could be illustrated in various ways. Let us first focus on a case involving politeness. Assume that I talk to some person whom I find utterly unsympathetic, although he gave me no specific reason for this assessment. Then I am allowed to believe that he is unsympathetic, but politeness forbids me telling him this straight away. So we seem to have a case in which I am permitted to believe what I am not permitted to assert.

Yet my concern here is only with moral norms; premise (1) means that if one may believe $[p]$, it is *morally* permissible to assert $[p]$. This does not imply an all-things-considered permission to assert $[p]$. Other norms, such as rules of politeness or Gricean conversational maxims, may forbid me to assert what is morally permissible, for instance, telling disagreeable people that they are disagreeable.

One could object that there is a moral obligation to observe the rules of politeness, so that I am not even morally allowed to assert something impolite. I am far from sure that a moral dimension is involved in obeying social norms such as rules of politeness; arguably, it is only for prudential reasons that one should comply with them. But there are other cases in which a moral impact can hardly be dismissed. For instance, if I promise you not to tell anyone the secret that you just told me, I am allowed to believe what you just told me—we can safely assume that I have no reason to distrust your testimony—but I appear to be morally obliged not to assert it, because in breaking the promise, I seem to violate not only some socio-linguistic norm, but also a moral one.

If this is right, then (1) is a defeasible moral norm. As it might be impossible to satisfy both, a moral norm obliging one to believe only what one is allowed to assert and a moral norm obliging one not to break a promise, we have to balance the conflicting norms against each other in order to arrive at an all-moral-things-considered precept for this specific situation. If we thereby arrive at the opinion that the norm not to break promises is stronger than (1) (or at least as strong as it), we are forced to revise (1) to

- (1*) If there are no overriding moral considerations, one must (morally): believe $[p]$
only if one is (morally) allowed to assert $[p]$.

Again, however, it is easy to see the harmlessness of the revision for my overall argument for moral evidentialism. For (1*) is only a reasonable alternative to (1) if we refer with ‘overriding moral considerations’ to moral factors that influence what one is allowed to assert but clearly not what one is allowed to believe. More precisely, conflicting norms that are to be filtered out by the ‘overriding moral considerations’ clause must not give us any instrumental, prudential, moral or other reason, however small, to believe differently than we otherwise would. Rules of politeness are a case in point, for they may have an impact on what we should not say to people, but do not affect what we should think about them. Promises are another case in point, as not even the promise to be silent about some matter gives us reason to revise our beliefs about that matter. Hence, whatever overriding moral considerations there might be, they have nothing to do with the formation or maintenance of beliefs.

The overriding moral considerations could even have an epistemic component. Assume that I know that you take me to be such a notorious liar that you will always come to believe the opposite of what I assert. Then I would deliberately misaddress your informational needs by asserting what I am allowed to believe. Arguably, this would give me an epistemically motivated moral obligation to refrain from making assertions when I would thereby risk to cause

you epistemic harm. This epistemic motivation, however, does not in any way affect what I am allowed to believe. For this reason, it is covered by ‘overriding moral considerations’ as well.²³

We are now in a position to state a modified and, if I am right, sound version of the original argument. If we substitute (1) with (1*) and (2) with (2**), the revised argument (which can easily be seen to be propositionally valid) is this:

- (1*) If there are no overriding moral considerations, one must (morally) : believe $[p]$ only if one is (morally) allowed to assert $[p]$. $[\neg OMC_1 \rightarrow (PBp \rightarrow PAp)]$
- (2**) If believing $[p]$ is not directly causally efficacious to a high degree and if there are no overriding moral considerations, one must (morally) : assert $[p]$ only if one is evidentially justified to believe $[p]$. $[\neg DCEp \wedge \neg OMC_2 \rightarrow (PAp \rightarrow J_eBp)]$
- (3) If, given there are no overriding moral considerations of the kind mentioned in (1*) or the kind mentioned in (2**), one must (morally) : believe $[p]$ only if one is evidentially justified to believe $[p]$, then one must (morally) : believe $[p]$ only if only if one is evidentially justified to believe $[p]$. $[[\neg OMC_1 \wedge \neg OMC_2] \rightarrow (PBp \rightarrow J_eBp)] \rightarrow (PBp \rightarrow J_eBp)$
- ∴ ME* Therefore, if believing $[p]$ is not directly causally efficacious to a high degree, one must (morally) : believe $[p]$ only if one is evidentially justified to believe $[p]$. $[\neg DCEp \rightarrow (PBp \rightarrow J_eBp)]$

The truth of the auxiliary premise (3) rests upon the logical fact that a proposition that is formally entailed by another one but independent from it holds *simpliciter*.²⁴ What remains to be shown to establish (3) is that ME* is independent from whether or not there are overriding moral considerations of the kind mentioned in (1*) or the kind mentioned in (2**). As outlined above, the overriding moral considerations mentioned in (1*) refer to moral factors that may affect what we are allowed to assert but not what we are allowed to believe, for instance, rules of politeness or demands to keep one’s promise to remain silent; and the overriding moral considerations mentioned in (2**) refer to moral factors that may affect what we are allowed to assert but not what we are evidentially justified to believe, for instance, norms that arguably accompany morally permissible lies. Hence, whether or not such norms are in place does not make any difference for what we are permitted or evidentially justified to believe. This includes that such norms do not have any impact on whether or not our beliefs should solely be based on the evidence we have.²⁵ Therefore, if an implication between what we are allowed to believe and what we are evidentially justified to believe is entailed by the absence of those overriding moral considerations, it holds quite generally. Thus, premise (3) is established.

It follows from this reasoning that moral considerations that may affect what we are allowed to believe must not be filtered out by the ‘overriding moral considerations’ clause. This is what makes the artist case in 3.3 so disturbing: it is a counterexample to both (2) and (2*), so that we have to substitute (2*) with (2**) and ME with ME*.

I do not regard ME* as merely stating a *pro tanto* obligation, an obligation that holds only as long as it is not overridden by some other obligation. The reason is that by claiming ME*, I hold something stronger, namely that the obligation in question is overridden only under those

²³I am grateful to Elizabeth Fricker for confronting me with this example.

²⁴Take a consequent to be independent from an antecedent if the truth value of the former does not depend on the truth value of the latter, e.g., if either both $p \rightarrow q$ and $\neg p \rightarrow q$ hold or both $p \rightarrow \neg q$ and $\neg p \rightarrow \neg q$ hold.

²⁵In other words (to take up the previous footnote), if $PBp \rightarrow J_eBp$ is true, both $(\neg OMC_1 \wedge \neg OMC_2) \rightarrow (PBp \rightarrow J_eBp)$ and $(\neg OMC_1 \wedge \neg OMC_2) \rightarrow (PBp \rightarrow J_eBp)$ are true, and if $PBp \rightarrow J_eBp$ is false, both $(\neg OMC_1 \wedge \neg OMC_2) \rightarrow \neg(PBp \rightarrow J_eBp)$ and $(\neg OMC_1 \wedge \neg OMC_2) \rightarrow \neg(PBp \rightarrow J_eBp)$ are true. Hence, independence holds.

very specific circumstances in which a belief is directly causally efficacious. To speak without further qualification of a *pro tanto* obligation to form or maintain evidentially justified beliefs would leave open under what circumstances the obligation is overridden, and I need not leave that open.²⁶

4.3 | Self-deception

If one is at t_1 aware that instrumental reasons (or whatever other reasons one may base a belief on but not an assertion) favour $[q]$ over $[p]$, while evidential reasons (or whatever reasons occur in the norm of assertion) favour $[p]$ over $[q]$, one could take means to bring oneself to think at some later point in time t_2 that, contrary to what one is aware of at t_1 , evidential reasons favour $[q]$ over $[p]$. At t_2 , one would then sustain a belief for evidential reasons that one aimed at adopting, at t_1 , for instrumental reasons. So, instrumental reasons may cause one to bring oneself to believe some proposition that one initially does not believe for evidential reasons and is therefore not allowed to assert.^{27, 28}

After she talked to her doctor, for example, Dolores is well aware that her evidential and instrumental reasons for believing that she will probably recover diverge. In this situation, she cannot simply believe against her evidence, because, as I pointed out, if one thinks that one has no reasons (in the wide sense of *reasons* sketched in footnote 20) for believing $[p]$, one automatically does not believe it. Dolores could, however, choose to take means to deceive herself about her evidence. These means can be more or less direct. More direct ones include visits to other doctors in the hope of getting a differing opinion and thereby a justification to disbelieve her original doctor's assessment, as well as attempts to convince herself that she, of all people, is far better suited than others to defeat the illness (maybe she had always been of unshakable health, or has done much sports during her whole life, or is a born fighter). More indirect means include hypnotists as well as science-fiction scenarios involving belief-inducing pills or brain surgeries, all of which cause her to believe that, according to her evidence, she is likely to survive. In short, although Dolores cannot, at will, bring herself to immediately come to believe that she will survive, she may succeed in getting herself to believe this at some point in time in the near future. So, it appears that she can believe that she will survive, while she is not allowed to assert it.

In response to such scenarios of self-deception, two things should be said. First, if we regard a single point in time, it is not the case that, at this point in time, a subject believes $[p]$, but could be blamed for asserting it. If my foregoing explanations are correct, one cannot, at t_1 —that is, before one started to deceive oneself—believe $[p]$ for reasons that do not warrant its assertion (instrumental reasons, say) or for no reasons at all. So, at t_1 , (1) is not violated. At t_2 , one has not only succeeded to bring oneself to believe $[p]$, one has also succeeded to bring oneself to believe that $[p]$ is supported by reasons that warrant assertion (evidential reasons, say). Given this belief, however, one cannot be blamed for asserting $[p]$, for this is correct according to all that one knows at t_2 . It then depends on our underlying account of evidence whether (1) is violated. If we assume evidential transparency, that is, that one's evidence is what one takes to be one's evidence, $[p]$ is indeed supported by one's evidence (although, of course, one has impoverished one's evidential situation considerably, and crucially, by deceiving oneself). So

²⁶Thanks to an anonymous reviewer for making me clarify this point.

²⁷This line of thought draws on Buckareff's amendment to Williams's famous but commonly considered to be refuted idea that due to the truth-directedness of belief, we cannot, for purely conceptual reasons, exercise direct doxastic control over acquiring doxastic states such as beliefs; see Buckareff (2014) and Williams (1973).

²⁸To be sure, there may also be cases of self-deception in which someone might be said to believe some proposition (say, that her child is very clever) and *at the same time* not believe it (as she knows that her evidence speaks against it). For the purpose of the present paper, I need not consider such cases.

(1) is satisfied. If we assume, on the other hand, that one can be mistaken about what one's evidence is, one's self-deception causes one to think that $[p]$ is supported by one's evidence, while in fact it is not. So in this case, one would indeed violate the norm of assertion if one were to assert what one believes, namely $[p]$; however, one could not be blamed for that, because one sincerely regards one's reasons for believing $[p]$ as warranting asserting it. So, at t_2 , one can at least not violate (1) on purpose.

A second and more important point is that one already violates the norm of assertion at the moment at which one begins to adopt measures for self-deception, because at that moment, one wilfully or carelessly²⁹ accepts that one will address the informational needs that other people might have at some later point in time in an inappropriate way. A self-deceiver is with regard to moral blameworthiness like someone who gets drunk before paying a visit to his or her ex-partner: as the person who starts drinking embraces the possibility that the drunken person he or she will come to be will act like a fool or a hooligan, this person can be blamed for provoking potential misbehaviour by getting drunk with the intention of paying the visit afterwards, even if the drunken person later on cannot help acting in such a way because of his or her drunkenness. Hence, the norm of assertion should be read to forbid attempts of self-deception from the outset in order to prevent people from getting themselves in situations in which they may not be able to comply with this norm or, if we assume evidential transparency, weaken their evidential situation.

Of course, one might regard certain reasons for self-deception as so urgent that they override compliance with the norm of assertion; but, as usual, the fact that rules are ignored on purpose by people who give more weight to other aims than to observing those rules is compatible with the rule-character of those rules.

4.4 | Retraction

A characteristic feature of assertions is their retractability. To retract an assertion is to make it invalid by taking it back. One may retract an assertion when one is no longer prepared to vindicate it; one is not forced to stay committed to everything one recently asserted. The reasons one has for such a retraction arguably need not be reasons for abandoning the belief that originally led one to make the assertion in the first place. One could retract an assertion because one thinks that one cannot defend it, but continue to believe its content (MacFarlane, 2010, pp. 83–84).

Assume, for example, that I remember having read much about a recent political decision, and consider it quite reasonable. Over lunch, I assert this to my colleagues, who immediately begin to attack my position, piling argument on argument to the point that this recent decision is disastrous. Unfortunately, I have forgotten what reasons originally brought me to adopt my position; although I feel confident that there are convincing replies to their objections, I am currently unable to produce them. As a consequence, I retract my assertion that the decision is reasonable, but continue to believe that it is.³⁰

The crucial aspect of this case (and of similar cases in which someone retracts an assertion without revising the respective belief) is that what the asserter accepts as relevant evidence differs from what the audience accepts as relevant evidence. In the example, I regard it as sufficient evidence for the proposition under debate that I remember having assessed it carefully and found it well-justified. Hence, I believe it throughout. My audience, on the other hand, is

²⁹While intentionalists concerning self-deception hold that self-deceptions normally stem from an intent to deceive oneself, non-intentionalists deny this by emphasising the role that biases and similar cognitive shortcomings play in our belief formation processes. See Bermúdez (2000) for an account of the debate as well as a nuanced argument in favour of intentionalism. If intentionalism is correct, deceiving oneself is always a wilful act; if non-intentionalism is right, the self-deceiver acts at least carelessly.

³⁰For two similar cases, see Goldberg (2015, p. 166).

unimpressed by my claim that there are convincing reasons that I am incapable of recalling, and accepts as evidence only those arguments that can be put on the table right now. The fact that I subsequently retract my assertion can then be explained on dialectical grounds: if I appeal to evidence that my audience does not admit in order to justify my assertion, I thereby run the risk of disqualifying me from the rest of the discussion, because, in a good debate, neither party should base their claims on evidence that the other calls into question. So, my answer to cases of retraction is that, as far as the moral reasons given by (1) are concerned, one is still allowed to assert what one continues to believe, because whatever justification one has for one's continuing belief suffices, by the argument given above, to justify the respective assertion as well, and the appearance that the assertion is not warranted anymore and should therefore be retracted is caused by dialectical norms of assertion that apply due to the contestedness of the underlying evidence. Were these specific dialectical norms absent—as they were, for instance, when I, later that afternoon, talk to other people, who have not yet questioned my evidence—my reasons for not asserting what I believe would have disappeared as well.

One might find some plausibility in the view that disregarding those dialectical norms is morally blameworthy, so that one is not even morally permitted to assert what the dialectical norms forbid. Even if such a view is truly defensible, one can, as in the case of conflicting norms (Section 4.2), substitute (1) by (1*) in order to preserve the argument for moral evidentialism.

Quite generally, cases in which one appears to be allowed to believe something but not allowed to assert it are cases in which the norm that forbids the assertion is either no moral norm at all or a moral norm that does not stem from epistemic considerations. Therefore, such cases do not refute (1*).

5 | CONCLUSION

The first and still most famous line of reasoning in favour of moral evidentialism is presented by Clifford in his influential paper *The Ethics of Belief* (Clifford, 1877/1999). According to Clifford, every belief that one maintains influences one's actions, and may easily influence them in a bad way if it is based on insufficient evidence, so that we are morally required³¹ to avoid beliefs based on such evidence in order to avoid those beliefs' negative consequences for our fellow human beings. My argument differs from Clifford's in that I am more specific about the harm that evidentially unjustified beliefs may cause: as there are close ties between the ethics of belief and assertion, evidentially unjustified beliefs put one at risk to address other person's informational needs inappropriately. As a consequence, my argument does not rely on Clifford's assumptions that every belief influences one's actions, and that this influence is generally worse in the case of an evidentially unjustified belief than in the case of an evidentially justified belief (as to the first, highly theoretical beliefs may not influence one's actions; as to the second, being overly optimistic may turn out helpful, as the case of Dolores, our deathly ill scientist, illustrates).

In addition, my argument does not rely on a presupposition that purely epistemic evidentialism has to take for granted in order to be worth considering, namely that we have an adequate motivation for believing what epistemic norms require us to believe. This presupposition can be doubted. After all, why should we believe what our evidence suggests, if not for prudential or moral reasons? Epistemic evidentialists need to give an answer to this question. I do not.

³¹Since Clifford does not distinguish moral from epistemic obligations (see Haack, 2001), it is unclear how he would spell out their relation.

ORCID

Marc Andree Weber  <https://orcid.org/0000-0001-5062-2404>

REFERENCES

- Adler, J. (2002) *Belief's own ethics*. Cambridge, MA: MIT Press.
- Alfano, M. (2013) *Character as moral fiction*. New York: Cambridge University Press.
- Basu, R. (2019a) Radical moral encroachment: the moral stakes of racist beliefs. *Philosophical Issues*, 29(1), 9–23.
- Basu, R. (2019b) What we epistemically owe to each other. *Philosophical Studies*, 176(4), 915–931.
- Basu, R. & Schroeder, M. (2019) Doxastic wrongdoing. In: Kim, B. & McGrath, M. (Eds.) *Pragmatic encroachment in epistemology*. New York: Routledge, pp. 181–205.
- Bermúdez, J.L. (2000) Self-deception, intentions, and contradictory beliefs. *Analysis*, 60(4), 309–319.
- Bird, A. (2007) Justified judging. *Philosophy and Phenomenological Research*, 74(1), 81–110.
- Boult, C. & Köhler, S. (2020) Epistemic judgement and motivation. *The Philosophical Quarterly*, 70(281), 738–758.
- Brogaard, B. (2014) Wide-scope requirements and the ethics of belief. In: Matheson, J. & Vitz, R. (Eds.) *The ethics of belief*. Oxford: Oxford University Press, pp. 130–145.
- Buckareff, A.A. (2014) Deciding to believe redux. In: Matheson, J. & Vitz, R. (Eds.) *The ethics of belief*. Oxford: Oxford University Press, pp. 33–50.
- Christensen, D. (2007) Epistemology of disagreement: the good news. *Philosophical Review*, 116(2), 187–217.
- Clifford, W.K. (1877/1999) The ethics of belief. In: Madigan, T. (Ed.) *The ethics of belief and other essays*. Amherst, NY: Prometheus, pp. 70–96.
- Conee, E. (1987) Evident, but rationally unacceptable. *Australasian Journal of Philosophy*, 65(3), 316–326.
- Feldman, R. (2007) Reasonable religious disagreements. In: Antony, L. (Ed.) *Philosophers without gods*. Oxford: Oxford University Press, pp. 194–214.
- Fritz, J. (2017) Pragmatic encroachment and moral encroachment. *Pacific Philosophical Quarterly*, 98(51), 643–661.
- Gardiner, G. (2018) Evidentialism and moral encroachment. In: McCain, K. (Ed.) *Believing in accordance with the evidence: new essays on Evidentialism*. Cham: Springer, pp. 169–195.
- Goldberg, S. (2015) *Assertion*. Oxford: Oxford University Press.
- Haack, S. (2001) ‘The ethics of belief’ reconsidered. In: Steup, M. (Ed.) *Knowledge, truth, and duty: essays on epistemic justification, responsibility, and virtue*. Oxford: Oxford University Press, pp. 21–33.
- Heil, J. (1983) Believing what one ought. *The Journal of Philosophy*, 80(11), 752–765.
- Hirvelä, J. (2023) The structure of moral encroachment. *Philosophical Studies*, 180, 1793–1812.
- Keller, S. (2004) Friendship and belief. *Philosophical Papers*, 33(3), 329–351.
- Kvanvig, J. (2009) Assertion, knowledge and lotteries. In: Pritchard, D. & Greenough, P. (Eds.) *Williamson on knowledge*. Oxford: Oxford University Press, pp. 140–160.
- MacFarlane, J. (2010) What is an assertion? In: Brown, J. & Cappelen, H. (Eds.) *Assertion: new philosophical essays*. Oxford: Oxford University Press, pp. 79–96.
- Maitra, I. & Weatherson, B. (2010) Assertion, knowledge, and action. *Philosophical Studies*, 149(1), 99–118.
- Marušić, B. (2013) Promising against the evidence. *Ethics*, 123(2), 297–317.
- Marušić, B. (2015) *Evidence & agency. Norms of belief for promising and resolving*. Oxford: Oxford University Press.
- Matheson, J. (2014) A puzzle about disagreement and rationality. *Social Epistemology Review and Reply Collective*, 3(4), 1–3.
- Matheson, J. (2015) *The epistemic significance of disagreement*. New York: Palgrave Macmillan.
- Meiland, J. (1980) What ought we to believe or the ethics of belief revisited. *American Philosophical Quarterly*, 17(1), 15–24.
- Mitova, V. (2011) Epistemic motivation: towards a metaethics of belief. In: Reisner, A. & Steglich-Petersen, A. (Eds.) *Reasons for belief*. Cambridge: Cambridge University Press, pp. 54–74.
- Navarro, P.E. & Rodríguez, J.L. (2014) *Deontic logic and legal systems*. New York: Cambridge University Press.
- Pace, M. (2011) The epistemic value of moral considerations: justification, moral encroachment, and James’ ‘Will to believe’. *Noûs*, 45(2), 239–268.
- Reisner, A. (2009) The possibility of pragmatic reasons for belief and the wrong kind of reasons problem. *Philosophical Studies*, 145(2), 257–272.
- Ryan, S. (2015) In defense of moral evidentialism. *Logos & Episteme*, 6(4), 405–427.
- Shah, N. (2006) A new argument for evidentialism. *The Philosophical Quarterly*, 56(225), 481–498.
- Stroud, S. (2006) Epistemic partiality in friendship. *Ethics*, 116(3), 498–524.
- Sutton, J. (2007) *Without justification*. Cambridge, MA: MIT Press.
- Way, J. (2016) Two arguments for evidentialism. *The Philosophical Quarterly*, 66(265), 805–818.
- White, R. (2005) Epistemic permissiveness. *Philosophical Perspectives*, 19(1), 445–459.

- Williams, B. (1973) Deciding to believe. In: *Problems of the self*. New York: Cambridge University Press, pp. 136–151.
- Williamson, T. (2000) *Knowledge and its limits*. Oxford: Oxford University Press.
- Wood, A. (2008) The duty to believe according to the evidence. *International Journal for Philosophy of Religion*, 63, 7–24.

How to cite this article: Weber, M.A. (2024) An argument for moral evidentialism. *Theoria*, 1–20. Available from: <https://doi.org/10.1111/theo.12557>