ORIGINAL ARTICLE

Disagreement without belief

Yonatan Shemmer[1] | **Graham Bex-Priestley[2]**

[1]Department of Philosophy, University of Sheffield, Sheffield, United Kingdom

[2]Inter-Disciplinary Ethics Applied Centre, School of Philosophy, Religion and History of Science, University of Leeds, Leeds, United Kingdom

Correspondence
Yonatan Shemmer, Department of Philosophy, University of Sheffield, 45 Victoria Street, S3 7QB, Sheffield, United Kingdom.
Email: Y.Shemmer@sheffield.ac.uk

Abstract

When theorising about disagreement, it is tempting to begin with a person's belief that p and ask what mental state one must have in order to disagree with it. This is the wrong way to go; the paper argues that people may also disagree with attitudes that are not beliefs. It then examines whether several existing theories of disagreement can account for this phenomenon. It argues that its own normative theory of disagreement gives the best account, and so, given that there is good reason to believe disagreement without belief is possible, there is good reason to think that disagreement itself is normative.

KEYWORDS
contextualism, disagreement, experimental philosophy, expressivism, non-cognitivism, normativity, metaethics, relativism, subjectivism

1 | INTRODUCTION

Here's a simple view of disagreement: two people disagree with each other just in case one believes a particular claim is true and the other believes it isn't. Some may want to expand this idea and say that people disagree when the joint content of their beliefs is inconsistent, thereby diagnosing disagreement between, for example, someone believing a disjunction and someone else believing the negation of each disjunct. Alternatively, others might prefer to focus on the impossibility of both beliefs being true instead of on logical inconsistency. What these views have in common is that according to all of them disagreement occurs only between beliefs.

What about other attitudes? Some philosophers believe there can be disagreements of taste.[1] One of the authors of this essay thinks mushrooms are delicious, and the other thinks they are not. Perhaps this is a disagreement about the tastiness of mushrooms. If it is, and if judgements of taste are not beliefs, then we have a counterexample. Other candidate attitudes include preferences (Angie prefers buying a house, while her wife prefers a flat), intentions ("Let's go dancing!" says Fred; "Let's not," says Ginger), credences (Cat has 0.3 credence in climate change, whereas Sajid has 0.5), and reactive attitudes (Mother is angry with me, but Father is forgiving).

Perhaps all these apparent counterexamples are really just proxies for disagreements in belief. One may suspect that Fred and Ginger disagree not due to their different intentions but in virtue of the fact that Fred *believes* the two of them should go dancing, whereas Ginger believes they should not, and maybe Cat *believes* the appropriate credence in climate change is 0.3, whereas Sajid believes it is not. In section 2 we argue against this suspicion. In section 3 we present our empirical survey of ordinary intuitions, which supports our position that the concept of disagreement applies directly to non-belief attitudes just as it does to beliefs. We then try to account for the phenomenon of disagreement without belief. In section 4 we examine several existing theories of disagreement that either naturally predict disagreements without belief or can be amended to do so, and argue they fail to match up with intuitions. Finally, in section 5 we argue our normative theory of disagreement (Bex-Priestley and Shemmer 2017) succeeds where the others have failed. This is the theory that people disagree if and only if they share a common project and, given the standards of their project, the divergence of their attitudes implies at least one of them has reason to change their attitude. Our dual conclusion is that disagreement without belief exists and that disagreement is normative.

2 | WHY WE CAN'T KEEP IT SIMPLE

Clear formulations of the simple view are rare. To the best of our understanding, this is in part because there are very few discussions of disagreement altogether, and in part because many philosophers took it for granted that something like the simple view is correct and therefore didn't see any need to provide an exact analysis of disagreement. On the simple view, disagreement is a relation that holds between beliefs or between people in virtue of their beliefs. As we explained above, the simple view is better understood as a family of views that analyse disagreement in terms of some or other impossibility of a combination of beliefs. On some variants the relation of disagreement holds when one belief is the negation of the other, on another when their content is inconsistent, and on yet another when they cannot both be true. Proponents of this family of views (henceforth we will speak of it as if it were a single view) are not oblivious to the suggestions that there might be disagreements whose grounds are "clashes" between non-belief attitudes. They argue, however, that the simple view can accommodate such potential counterexamples. According to them, all such purported cases of disagreement in non-belief attitudes are to be understood as cases of derivative disagreement (Frances and Matheson 2019, esp. § 2).[2] Two people with different tastes disagree derivatively only if the divergence between their tastes has its origins in disagreement in beliefs about what the right taste is. In some such cases in which belief disagreement leads to divergence in taste, we can, speaking loosely, say that two people disagree in taste. However, strictly speaking, claim the proponents of the simple view, the only disagreement is belief disagreement.

In the current section we present two arguments against the analysis of purported counterexamples suggested by the simple view. The first argument appeals to ordinary usage. The second argument appeals to group disagreement.

---

[1]See, e.g., Kölbel 2004; Sundell 2011; MacFarlane 2014; Marques and García-Carpintero 2014; Ferrari 2015.

[2]Frances and Matheson focus on analysing disagreement in intention as derivative, but the strategy can be extended to all other types of non-belief attitudes.

2.1 | The argument from ordinary usage: Three examples

Consider the episode from Twain's *Huckleberry Finn* in which Tom and Huck haggle over the method in which they will free Jim. At this point in the story Tom and Huck have already decided to do so together. Their intentions regarding the following steps, however, are not aligned. Huck intends to leave with Jim right away. Tom refuses and insists that Jim must first have the "prisoner's escape experience." We are not sure how best to interpret Twain's narrative. Are Tom and Huck in disagreement about the "right way" to free a runaway slave? Maybe! If so, the simple view applies nicely to this case. But we ask the reader to leave aside for a moment the best interpretation of Twain's novel and to follow us in imagining a possible variant of the original story. In this variant Tom and Huck, two children aged twelve and thirteen, respectively, have no particular view about the right or just or most heroic or most beautiful way to free runaway slaves. Their desires, however, diverge. This divergence in their desires is driven and explained by the different fantasy worlds that they have imagined inhabiting in past playtimes. Tom's imagination is often immersed in tales of "arduous and elongated prison escapes" about which he reads in his children's books. Huck has his imagination captivated by the glory of the moment of gaining freedom. These different background fantasies, and their resulting different desires, lead them to form different intentions—intentions that they quickly discover differ from one another.

We claim that the following expressions are natural ways of describing the state that the children are in: "Tom and Huck disagree about how to free Jim," and "Tom and Huck disagree in their intentions." Importantly, while it is natural to describe the case as one of disagreement in action or intention, the diverging intentions are not grounded in disagreeing beliefs. Indeed, in our variant of the story, they couldn't be grounded in disagreeing beliefs, since the requisite beliefs are not held by the two children.

Consider another example. Jane and Joe are friends. They dine out together every Thursday evening. Tonight is no exception. Jane picks up Joe in her car, and the two discuss where to go. But tonight *they cannot agree.* Jane insists on Chinese food, Joe will only go for pizza. It may be true of some other Janes and Joes that their insistence is grounded in different beliefs about the right place to eat dinner at this time of year, or maybe in different beliefs about the right place for them to eat dinner given the places they have already been to on previous Thursdays. But not our Jane and Joe. They each simply *feel like* having a different dinner. Can we imagine such a Jane and Joe? Yes we can! And is it natural to describe them as disagreeing about where to have dinner? Absolutely!

Finally, consider a third case, a more general version of the previous two. Tim and Tammy are getting a divorce. When asked why, Tammy says, "For the past five years we've disagreed about everything. He wants a house in the country and I in the city, he intends to have only one child and I want three, he watches sports and I like dramas." How shall we fill in the details of the case? In one version of the story Tim and Tammy form intentions, develop tastes, and nurture desires only on the basis of beliefs about the best intentions, the right tastes, and the most virtuous desires. But this is only one version, and we imagine it's not particularly common among actual divorcees. On different versions the desires, intentions, and tastes are not the result of any belief. In these versions of the story, versions that are not thereby made less likely, the intentions, desires, and tastes are not even accompanied by "matching" beliefs. Must Tim really believe that sports are a better use of one's, or even his own, time than dramas? No, he simply likes watching sports. And nevertheless, Tammy's description of their marriage—*we've disagreed about everything*—strikes us as natural.

A defender of the simple view may offer two responses. First, she might say, for some or all of these attitudes, humans normally do have an underlying belief that their attitude is the correct one. Perhaps intentions are formed because we believe we ought to do the thing intended. Some philosophers do link intentions tightly with such beliefs (Gibbard 2003; Wedgwood 2007), but it is a controversial commitment to make (especially in the case of other attitudes, such as

desires and tastes), and we offer reasons to doubt it in the following subsection. Second, the defender may resist the intuition and insist there can be no genuine disagreement without conflicting beliefs; Tammy's description is not strictly true, however natural it may seem. While we accept it is possible for the proponent of the simple view to bite the bullet, it is a theoretical disadvantage, and we should prima facie prefer other theories that accommodate common intuitions. We will put more pressure on this response in section 3.

2.2 | The argument from group intentions

We start with a story. Our story has a non-fictional background: The North Atlantic Council is responsible for NATO's operations. Its decisions must be unanimous. Members of the council represent the position of the member states. In March 1999 the NAC decided on operation Allied Forces, which involved air strikes against the Serbian (officially Yugoslavian Army) forces.

The rest of the story is fictional: It is February 1999. Half of the British members of Parliament think that the best thing for NATO to do is to bomb the Serbian forces. The other half of the members have no view on the matter but still vote in favour of the bombing because they think that voting otherwise will hurt them in the next election. Half the members of the French National Assembly think the best thing for NATO to do is not to bomb the Serbian forces. The others have no view on the matter but still vote against bombing because they think that voting otherwise will hurt them in the next election. Given the story told so far, the following could be the case.[3]

1. The British Parliament has a shared intention to support the bombing. The French Parliament has a shared intention to reject the bombing.
2. The respective shared intentions of the two parliaments are not grounded in disagreeing beliefs.

Whether # 1 is true depends on two further assumptions. First, that there are indeed such states as group intentions. Second, that the most successful analysis of such states does not require shared beliefs that bombing is the best thing to do or the best thing for any particular individual or group. Of course, the fact that all the members of a parliament vote for the bombing does not ensure that they share an intention to bomb. We assume for the sake of argument that all the other conditions of the best account of shared intentions are being satisfied. One might wonder, however, whether it is indeed the case that the best account of shared intention to φ does not require a shared belief that φ is the best thing to do (or the best thing for the group to do). Defending this claim would take more space than we have on the current occasion. Instead we point out that, for example, Tuomela and Miller (1988), Searle (1990), Gilbert (2006), Tuomela (2007), and Bratman (2014) all present accounts according to which a joint belief that φ is the best thing, or the best thing for the group, is not necessary for a joint intention to φ.

The second claim (# 2) is supported by the fact that each parliament is completely divided on the beliefs that ground their intentions (Sunstein 2007). So even if we could trace (and maybe even conceptually connect) the parliament's shared intentions to the beliefs of its individual members, we cannot connect them in the requisite way to the shared beliefs of the group.[4] If we accept # 1 and # 2, it seems plausible to conclude that:

---

[3]The next two bits of the story might be disputed. We therefore present them as explicit assumptions with a supporting justification.

[4]We don't assume any particular view of "shared beliefs" and certainly do not assume that shared beliefs are determined by the majority of the beliefs of individual members. We merely claim that in the scenario we have described it is obvious that there is no shared belief one way or another.

3. Britain and France disagree on whether to bomb the Serbian forces.

And crucially,

4. Given the structure of the individual beliefs of the members of each parliament, it is impossible to correctly attribute the relevant group beliefs to the two parliaments.

Conclusion # 3 reflects common usage of the word "disagreement." In this respect the argument here is not different to the argument from intuition that we presented in section 2.1. What is special about the argument from group intention is the addition of conclusion # 4. A philosopher with particularly ardent Socratic inclinations might insist that, *pace* our examples in section 2.1, an individual could not form an intention to φ unless he or she really believed that φ is the best possible option. But even a philosopher with such radical views couldn't attribute the relevant group beliefs to the two parliaments in the story above. The distribution of beliefs among the individual members of the parliaments are such that it would be impossible for them to believe as a group that bombing is right or that bombing is wrong, or even right for the British or wrong for the French.

3 | EMPIRICAL RESULTS: WHAT SAY THE FOLK?

The scenarios we described in section 2 are psychologically possible, maybe even common. It is thus hard to object to the arguments presented in that section on the ground that they are based on implausible scenarios. Proponents of the simple view might instead object that our intuition about the proper application of the term "disagreement" to these scenarios is misguided. They might insist that it is unnatural to regard cases with no belief disagreement as cases of disagreement.

Spurred by such an objection, we decided to test whether an application of the term "disagreement" to clashes in non-belief attitudes was indeed natural. We presented two hundred subjects with a scenario of non-belief disagreement. We then checked whether they were comfortable with the use of the term "disagreement" to describe that scenario. The full account of the experimental setup and the full results of that experiment are to appear in a separate publication. In the next few paragraphs we merely describe the essentials of the experiment and its major results. We used the following vignette:

> Buba and Kiki are friends. They live in a remote community in which being indecisive is seen as a terrible vice. Members of this community have developed strategies in order not to be indecisive. Whenever they face a choice between options that seem equally good to them, they pick one of the options arbitrarily and then they form an intention to act on that option—without changing their view that the options are equally good.

> Buba and Kiki have decided to spend the evening together. The options available to them are to go to a movie or to go to a restaurant. Since, to both of them, these options seem like equally good ways to spend the evening together (they both think that there isn't a best way to spend the evening together), each one of them resorts to the strategy for avoiding indecisiveness. Buba arbitrarily forms the intention to go to a movie and Kiki arbitrarily forms the intention to go to a restaurant. "I am only willing to go to a movie," says Buba. "I am only willing to go to a restaurant," says Kiki.

The story is complex, and you might wonder why we needed the additional details. Why couldn't we have simply presented a case of a clash between non-belief attitudes, say, intentions, and left it at that? The answer is that the experimental subjects might assume that such clashes were accompanied by, or were the result of, a background belief disagreement. If they did assume this, and if they would then describe the vignette as a case of disagreement, proponents of the simple view would fairly claim that it is in virtue of the subjects' assumption about the background belief disagreement that they applied the term. The point of the additional details is to ensure that subjects understand the vignette as a case in which there is no background belief disagreement. In other words, we wanted to ensure that if the term "disagreement" were applied to the vignette it would be only in virtue of the clash between non-belief attitudes.

The setup: we presented subjects with the vignette and asked them whether Buba and Kiki disagree. Since we further worried that subjects might assume that as a result of forming their respective intentions Buba and Kiki also changed their view on the best way to spend the evening together, we controlled for that and excluded those subjects who made this assumption. In fact, very few did. After also controlling for other misunderstandings, we were left with a group of "relevant respondents." The results were staggering: of the relevant respondents, *80 percent thought that Buba and Kiki disagree*.

We also worried that respondents might think that Buba and Kiki's plans are in *some* kind of tension short of disagreement, but that since the respondents were given no way of expressing this thought, they would be "pushed" by the experimental design to describe that tension as a case of disagreement. We therefore ran a different variant of the experiment to control for this possibility as well. The results stayed virtually the same.

The empirical results indicate that our own intuitions about the application of the term "disagreement" are widespread. Yet an objector might argue that even if we have successfully shown that the folk believe that the concept of disagreement can be correctly applied to clashes of non-belief attitudes, this doesn't establish that they're right. The folk's understanding of a concept might be incompatible with established scientific facts, or incoherent or confused in some other way.

The objector is right in principle. A philosophical and/or scientific analysis might very well establish that the folk are confused about the nature of a certain concept. We contend, however, that mere possibility does not yet establish a ground for rejecting the popular understanding. The burden is on the objector who wants to reject the popular understanding of a concept to show why it must be rejected. The idea that disagreement applies to clashes in non-belief attitudes is certainly in no tension with scientific truth. Therefore, to argue that popular understanding of the concept of disagreement is misguided one would have to show that popular understanding is somehow confused or incoherent. One obvious way to do so would be to show that there is no unified theory of disagreement that can accommodate both belief and non-belief disagreement. In fact, in "A Normative Theory of Disagreement" (Bex-Priestley and Shemmer 2017) we initially argue along this very line; we show that certain theories of disagreement are inadequate because they treat these two types of disagreement as having little in common. But not all theories of disagreement are thus disunified. Crucially our own theory (presented in section 5) treats belief and non-belief disagreement as a single phenomenon and thus blocks the main form that the objection to the popular understanding might take.

A more radical objection is that folk usage, even when it is not incoherent or unscientific, should not be the basis for deciding which theory is correct. (We thank an anonymous reviewer for raising this point.) As Moore would have put it, our task as philosophers is not lexicographical. Philosophers should stop worrying about describing "folk" concepts and use whichever concept best suits their purposes. We do think there are strong philosophical reasons to accept our theory that are independent of folk usage, such as the fact (or so we contend) that it is the best theory for making sense of the notion of disagreement as it is used by expressivists and dialetheists. There are, however, lines of argument against certain philosophical positions

that rest on the assumptions that (a) folk concepts *are* important and (b) the folk concept is something like the simple view. In metaethics, for instance, some object to expressivism, subjectivism, contextualism, and relativism on the basis that they fail to accord with common sense understandings of disagreement, where these "common sense understandings" are implicitly assumed to be the simple view. Since we aim to oppose these lines of argument, we can legitimately challenge their assumptions about the folk position. While our experiment is not conclusive, we hope to have at least shifted the burden of proof to our opponents. It should no longer be taken for granted that the folk rule out disagreement between non-belief attitudes.

In sections 2 and 3 we advanced a variety of arguments against the simple view. We used case studies to elicit the intuition that clashes in desires, intentions, tastes, and other non-belief attitudes could be cases of disagreement even when they are not grounded in belief disagreement, and we presented empirical evidence that this intuition is widespread. One might still wish to hold on to the simple view on the grounds that there are no good theoretical alternatives. In the next section we discuss and reject some existing theoretical alternatives. In section 5 we present our own theory of disagreement. Our theory demonstrates that non-belief disagreement can be well systematized. Importantly. it also demonstrates that belief and non-belief disagreements can be understood as two instances of a single phenomenon.

4 | RIVAL THEORIES OF DISAGREEMENT

Unlike the simple view, other theories are amenable to disagreements without belief. In this section we consider five of them. This, of course, doesn't cover every alternative to the simple view.[5] The weaknesses exposed will, however, serve to motivate our own theory.

4.1 | Self-identification

The idea that disagreement must be recognised by the disagreeing parties does not appear in the literature but was suggested to us in conversation by several philosophers.[6] In public use it sometimes seems that the word "disagreement" is more readily applied to participants who are *unsettled* or in some kind of tension with each other. This is especially true with the fuller phrase "having a disagreement." One formulation of the idea is that people disagree if and only if they believe they do. This can accommodate disagreements without belief and is quite flexible in doing so. Suppose Tamsin prefers stoner metal to thrash metal, and Arne prefers thrash to stoner. If they take themselves to be in disagreement, perhaps because their preferences cause some tension and argument between them, then they are indeed in disagreement. The self-identification view thus allows for a disagreement in preference. It's also flexible because if Tamsin and Arne take themselves to have merely *different* rather than disagreeing preferences, then they don't disagree.

An initial problem with the view as formulated is that the concept of disagreement appears on both sides of the biconditional. A second problem is that it does not match our intuitions. On the one hand, it seems possible for people to mistakenly think they disagree when they don't. A case of "misagreement" could be when one person asserts that Tommy is a lovely boy and another person "dissents," not realising that the two of them have different boys named Tommy in mind. On the other hand, it seems possible for people to disagree even if they don't realise it. We could imagine an alteration of the previous example where the two people are

---

[5]A notable omission is Ridge (2013; 2014, chap. 6). We argue against Ridge's view (and include more detailed arguments against Stevenson and Gibbard) in Bex-Priestley and Shemmer 2017.

[6]Jimmy Lenman, for example, suggested that the folk conception might be something like this.

referring to the same boy but each thinks the other person has a different boy named Tommy in mind. We can also imagine small children disagreeing about where the toy is hidden before they have learned the concept of disagreement. While this theory can accommodate disagreement without belief, it does so at the cost of these counterintuitive consequences. It's worth looking at other options.

4.2 | Stevensonian views

Charles Stevenson (1944, 3) claimed there could be disagreements in desire as well as belief. Whereas conflicting beliefs cannot both be true, conflicting desires cannot both be satisfied, and this is what grounds disagreement. The problems with this view have been exposed in detail elsewhere (Ridge 2013; 2014; Bex-Priestley and Shemmer 2017), but since this theory has been influential for several philosophers, including Simon Blackburn (1998, 69) and Stephen Finlay (2014, chap. 8; 2017, 193), it is worth briefly rehearsing them.

First of all, as Mike Ridge (2013, 46; 2014, 171) notes, it seems as though "disagreement" picks out a unified kind. Stevenson has trouble explaining this unity. If inconsistency is the unifying feature, why don't inconsistent fantasies count as disagreement? Defenders of Stevenson may suggest that the proponents of the simple view are companions in guilt, since they also need to explain why inconsistency matters in beliefs but not in other attitudes. These proponents, however, do have an explanation of what unifies disagreement-apt attitudes: *they aim at truth*. It doesn't matter if people's desires, fantasies, suppositions, and so on are jointly inconsistent, because those mental states weren't aiming at truth in the first place. Stevenson's defenders are therefore unique in owing us an explanation for what unites the attitudes that are disagreement apt from those that are not.

The other problem is its counterintuitive results. Stevensonian views come in two varieties, and both get the extension wrong. In the weak version, mere inconsistency in desire is sufficient for disagreement. This lets in too many cases. It would mean chess opponents who each desire to win would count as being in disagreement, but this is a very unnatural consequence. It is unclear what they disagree about; they just want to win. Similar things can be said about any conflict, such as independent burglars trying to steal the same diamond or countries fighting over land.

In the strong version an extra condition is added: "at least one of them has a motive for altering or calling into question the attitude of the other" (Stevenson 1944, 3; see also Stevenson 1963, 1). This correctly rules out the chess opponents, and it links with the idea in the previous subsection that disagreement indicates a kind of unsettled situation. As we indicate in section 5, our own view is that this sense of an unsettled situation reflects the suspicion that one of the disagreeing parties has a reason to change their mind. Yet the "motive" condition provided by Stevenson is far too strong. We may disagree with our dying relative about the existence of an afterlife even if we are not motivated in the slightest to alter or call into question our relative's attitude. This is a benevolent case of lacking the motive condition, but there are also cases of malevolent deception. Consider Ar-Pharazôn and Sauron in *The Silmarillion* (Tolkien 1977, 330–36). Ar-Pharazôn, due to Sauron's cunning, believes it would be good for Númenor to attack Valinor; Sauron knows it would be their doom. Again, they intuitively disagree. But Sauron wants the Númenóreans to attack and be destroyed, so he has no motive for altering the current attitude of Ar-Pharazôn. The take-home lesson is that disagreeing parties needn't have any wish to change each other's mind. In light of these problems, we should reject Stevensonian views.

4.3 | Disagreement by rejection

The theory of disagreement by rejection can be found in works by Allan Gibbard (2003, chap. 14) and Graham Priest (2006, chap. 6).[7] Priest discusses only beliefs and Gibbard only beliefs and plans, but the idea can in principle be extended to any attitude. According to this theory, to disagree with someone is to reject one of their attitudes. We can think of the rejection of an attitude as a resolution not to adopt it. "To reject something is to refuse to believe it: if it is in one's belief box one takes it out, but whether or not it was in there before, one resolves to keep it out" (Priest 2006, 103).[8] For example, I may consider the evidence and decide to believe the butler did it, to believe the butler didn't do it, or to reject forming either of these beliefs. If I end up believing the butler did it and you reject that same belief then, whether or not you believe the butler *didn't* do it, we disagree.

Extending this beyond belief, suppose I then decide to prosecute the butler and you decide to reject this plan (Priest 2006, 103).[9] This counts as a disagreement. We might call it a disagreement in plan or a disagreement about what to do. Here are some other possible disagreements: I place 0.8 credence in the butler's guilt, and you reject this credence; I feel angry with the butler, and you reject this anger; and I find the butler's sophistry distasteful, while you reject this distaste. Each of these could be contested, of course. One may think it's not really possible to keep something out of one's "taste box," for instance. It's not clear how to determine which mental states are rejectable and which aren't, but since the most intuitively rejectable states are beliefs and intentions, our argument against this theory will focus on disagreements of intent. The argument appears in the next subsection because it applies also to the following theory of disagreement.

4.4 | Interpersonal incoherence

Alex Worsnip believes that "[y]ou and I disagree in the wide sense if and only if I hold some (positive) attitude or attitudes A1, and you hold some (positive) attitude or attitudes A2, such that the combination of A1 and A2 in a single individual would produce an incoherence (that was not already present in either A1 or A2 taken individually)" (2019, 259).[10] Flat-Earther Frank disagrees with Spherical-Earther Sophie because if one person were to simultaneously share Frank's belief that Earth is flat and share Sophie's belief that Earth is spherical, this person would become incoherent. Different credences in the same proposition will also be counted as disagreements. Not only can Worsnip's theory accommodate various disagreements without belief, it can also accommodate disagreements in different kinds of attitudes at the same time: Becky, who intends to make a time machine, disagrees with Kristov, who believes time travel is impossible (assuming it's incoherent to intend what one believes is impossible).

---

[7]It might not be fair to include Gibbard. While he does discuss disagreement in terms of rejection, he also says he takes disagreement to be primitive (2003, 74; 2012, 273–74). If disagreement is primitive, then perhaps there can be disagreements without belief, and perhaps we could stipulate that the extension matches up with our intuitions. We don't have much to say about this idea, because we are trying to come up with a theory of disagreement rather than leaving it as a brute unexplained fact.

[8]As you can see, Priest talks only of rejecting beliefs.

[9]More specifically, you must decide to reject the following *conditional* plan: to prosecute the butler *were you in my exact situation*. Merely rejecting the plan to prosecute doesn't mean you disagree: maybe you don't have the authority to do so yourself, but you agree with the plan to do so in my very circumstances.

[10]See Dreier 2009 for similar ideas, as well as Baker and Woods 2015 on the possible "discordance" of attitudes taken together.

The objection to Worsnip's theory, which as foreshadowed earlier applies also to theories of disagreement by rejection, is that it sees disagreement where intuitively there is none. To illustrate, consider a variation of the scenario we gave to participants in the experiment described in section 3, a variation we'll call "BK2" (Buba and Kiki 2). Buba and Kiki are again deciding what to do, but this time they are happy to act *separately*. They believe all the outcomes are equally good—both of them seeing a film, both dining, or one doing one while the other does the other—and so they form their intentions arbitrarily and reject the other options. Buba ends up intending to see a film, and Kiki intending to dine. They enact their individual plans. Do Buba and Kiki disagree? It seems to us that in this case they do not. They may have had different plans, but there is nothing to suggest one of them was wrong, and there was no tension between them to resolve. Yet according to the theories of disagreement by rejection and interpersonal incoherence, Buba and Kiki *were* in disagreement. Buba rejected the option Kiki planned to do, and vice versa, and it would be incoherent to adopt both of their plans (assuming one cannot simultaneously see a film and dine in this context). In short and speaking a little loosely, contrary to disagreement by rejection I might acceptably have an attitude myself while you, equally acceptably, reject that attitude yourself; contrary to disagreement by interpersonal incoherence, the two of us might have attitudes that are perfectly acceptable individually and collectively but would be incoherent in the same head.

Worsnip foresees this worry. His response (2019, 263–64) is to distinguish between centred plans and non-centred plans and to deny that centred plans exist. According to Worsnip, Jenny does not plan *to leave the party*; she plans *for Jenny to leave the party*. This is a controversial move. Many philosophers want to accept centred objects for attitudes, such as *de se* beliefs. If Serena also plans to leave the party, we naturally describe Jenny and Serena as having the same plan. Nevertheless, if we grant that centred plans do not exist, then Buba's plan is *for Buba to see a film*, whereas Kiki's plan is *for Kiki to dine*, both of which are coherent in the same head. What does it mean for someone who isn't Kiki to form the plan *for Kiki to dine*? It is not to plan to make Kiki dine, or to persuade her to dine, or to do whatever is in one's power to get her to dine. It is to form a conditional plan: *upon finding oneself in exactly Kiki's situation, dine*.

One reason to look for another account is that Worsnip's theory has this controversial commitment. It would be nice not to have a stake in the debate on centred propositions for attitudes. More importantly, though, it's not clear that it will do the trick. Let us strengthen BK2 such that Buba and Kiki—knowing they are in similar enough circumstances—formulate their respective plans *for both of their situations*. Buba, having decided there's no right or wrong thing to do, arbitrarily forms the dual intention *for Buba to see a film* and *for Kiki to see a film*. Were someone to hypnotise Buba into thinking he is Kiki right now, Buba would stick with the plan to see a film. Likewise, Kiki plans to dine whether she's in Buba's shoes or her own. This doesn't seem particularly strange. Neither Buba nor Kiki has a preference for one option over the other, so the arbitrary intention serves well enough in either situation. We can now formulate the same argument. Buba and Kiki don't seem to disagree with each other, because they both agree that neither plan in either of their situations is wrong and both of them can carry out their plans no matter what. There is no tension between either of them, and we see no problem with their attitudes. Their different plans *for what to do as Buba* would, however, be incoherent in the same head. This shows that identifying disagreement by interpersonal incoherence, as well as by rejection, causes us to see disagreements where there are none.

4.5 | Objective standards

We believe that the missing ingredient so far has been normativity.[11] The fact that one person rejects a personal plan that another person has, or that two people's minds would not form a coherent single mind, does not yield the kind of tension that is constitutive of disagreement. The final theory of disagreement we now consider before offering our own comes from Peter Railton, where the normativity is explicit. His theory has not been published, but he has suggested to us that a view such as this could be formulated on the basis of his view of desire, value, and learning presented in his 2012 APA Presidential Address and in his 2018 Locke Lectures.

The underlying idea is that beliefs are not the only attitudes with objective standards of rationality and correctness. Desires also have such objective standards that guide their formation. In some cases of drug abuse, for example, the desire to continue taking drugs is the incorrect desire to have. In cases where an agent has the correct goal to φ and the correct belief that ψ-ing is the only way to φ, it would be incorrect for them to intend not to ψ. According to Railton, disagreement occurs only when a mistake—a failure to meet these standards—is guaranteed: people disagree in attitude if and only if there is an objective standard of correctness for that attitude and, given this standard, it is impossible for all the attitudes to simultaneously avoid being incorrect. The dependence on a standard of correctness provides the missing normative element.

This makes a lot of sense to us. It parallels how we think of disagreement in belief: if two people disagree in their beliefs, at least one of them must be wrong. It also gets the right answer in the case of BK2, which plagued the theories of disagreement by rejection and interpersonal incoherence. Assuming that Buba and Kiki are correct in their assessment that the relevant options are equally good, neither person formed an incorrect intention. Both intentions are rationally permissible, and Railton rightly classifies this as a case of no disagreement. Unfortunately, it fails the *original* Buba and Kiki case, where they are going to act together. Assuming again that there is no objectively better option, neither Buba nor Kiki formed an objectively incorrect intention. (Who was wrong: Buba or Kiki?) On Railton's view the original Buba and Kiki case, which we have argued is a case of disagreement, fails to count.

Railton may reply that a mistake *was* made by Buba and Kiki in the original case. When you're going to act together it's incorrect to arbitrarily form your intentions individually, otherwise you might end up in their tragic situation. You should flip a coin together or something, not make plans behind closed doors. We agree that this is good advice. An incorrect *procedure*, however, is not sufficient to ground disagreement. Indeed, incorrect procedures may well lead to complete agreement. Buba and Kiki could have individually formed the same plan, and it would be odd to label this as a disagreement despite their mistaken procedure. This reply won't work.

What lessons have we learned? The ways in which the first four theories of disagreement fail to match up to intuitions suggest to us that a normative condition is part of our concept of disagreement. In this subsection we have seen, however, that objective standards are not enough to capture all cases of disagreement. It is possible to disagree with people in situations where we are not bound by objective standards of correctness or in situations where such standards give us sufficient leeway to permissibly form each of the disagreement-apt attitudes. In 2017 we published our own normative theory of disagreement as a solution to some problems

---

[11]Worsnip does say that incoherence is normative, but his understanding of it is very thin: "[T]his philosophical role for coherence is entirely independent of the controversy about whether it's valuable to be coherent, or whether we have any reason to be coherent" (2019, 266). We're interested in a thicker notion of normativity that does involve reasons. Furthermore, even if it turns out that it is valuable to be coherent, this does not mean it is valuable for us to have minds that would be coherent if they merged. Given that no mind mergers happen, it seems unlikely that I have reason to form the plans you form or to reject the plans you reject (unless we'd get in each other's way).

for expressivism in metaethics. We offer this theory here as a general account of disagreement with and without belief. Our theory both introduces the normative condition missing in the preceding theories and allows for disagreement between people in virtue of attitudes that are not bound by any objective standards.

5 | A NORMATIVE THEORY OF DISAGREEMENT

We believe that people disagree only when at least one person has reason to change their mind. These reasons do not need to be grounded in objective standards related to the correctness of attitudes. The formulation of our theory (Bex-Priestley and Shemmer 2017, 198) is as follows. A and B disagree if and only if:

1. A has attitude *a*, B has attitude *b*, and *a* and *b* are different.
2. A and B share a common project with certain standards for the formation and retention of attitudes. (Whether people are involved in a common project may itself be a normative matter. We take no stand on this issue.)
3. Given the standards imposed by their common project, the divergence of *a* and *b* implies that at least one of A and B has reason to change their attitude.

The "common project" we refer to can range from purposefully undertaking a specific activity together to merely having the same goal. Some examples will help illustrate the theory. Alice believes that vaccines cause autism, and Bernie believes they don't. Even if they've never met, they share a common project of believing the truth. Given the standards for beliefs, it's clear that at least one of them—whoever is wrong—has reason to change their attitude. Importantly, this applies even if they have both reasoned correctly from their evidence. Subjective reasons can come apart from objective reasons, and the fact that a proposition is false is an objective reason not to believe it even if we would not be at fault for doing so. (For more details on this issue and a discussion of potential objections, see Bex-Priestley and Shemmer 2017, 204 and § 6.) The three conditions are met. Alice and Bernie disagree about whether vaccines cause autism.

Amelia prefers vanilla ice cream, and Brooke prefers chocolate. Perhaps they're not really engaged in any common project, or perhaps they both have the common goal of enjoying ice cream. Either way, there is no implication that Amelia or Brooke has any reason to change her preference. *Vive la différence*. The three conditions are not met. Amelia and Brooke do not disagree.

Andy plans to paint the fence red all over, and Bianca plans to paint it green. Given the divergence in their plans and their common project of painting the same fence, there is reason for at least one of them to change their plan. Note that this reason does not come from one of the colours being objectively incorrect. In this case the reason comes from the fact that their diverging intentions frustrate their shared goal. The three conditions are met. Andy and Bianca disagree about what colour to paint the fence. Now, we could re-describe the case so that Andy and Bianca do not have the usual goals when it comes to painting fences. Maybe they are involved in a competition where the winner is whoever covers more of the fence in their own colour. In this case they no longer disagree. Even though they cannot both succeed, the standards of their competition along with Andy and Bianca's divergence in plan do not imply any reason for either person to abandon their intention. We think this is the right result. What separates disagreeing intentions from non-disagreeing intentions isn't merely their incompatibility, as the weak Stevensonian views have it, but the implication of reasons to change.

According to our theory, then, the same mental states may constitute disagreement in some contexts but not in others. This is a feature, not a bug. It is why we can correctly classify both

of the Buba and Kiki cases. In the original case, the one used in our experiment, Buba and Kiki are involved in a common project. The goal is to spend the evening together. Given this project, and given their diverging intentions, there is reason for at least one of them to switch plans so the two of them are in alignment. The three conditions are met, and the common intuition that Buba and Kiki disagree is vindicated. In BK2 the two mental states in question are the same as before, but since Buba and Kiki no longer share the project of spending the evening together, there is now no implied reason for either individual to alter their intention. The normative landscape has shifted. Thus, our theory explains why our intuitions about whether or not Buba and Kiki disagree change depending on the background conditions and not simply on the mental states in question.

Our theory can accommodate cases of mistaken agreement and mistaken disagreement. The divergence between someone's belief that Tommy-1 is a good boy and someone else's belief that Tommy-2 is not a good boy does not indicate any reason for either person to change their mind. The existence or non-existence of disagreement does not depend on successfully identifying it as such. It also does not depend on either party having any particular motive. We disagree with our spiritual relative about the existence of an afterlife because, given our project of having true beliefs (or at least avoiding false ones), there *is* an indication that one of us has a reason to change their mind: if there's an afterlife it's us, but if there's not, it's our relative.

Readers may have independent objections to our theory of disagreement. We cover some of them in our 2017 paper (esp. pp. 201–26). In the present paper we have focused more on the ways in which theories can accommodate disagreement without belief and we have argued that ours has the best chance of getting the extension right. Naturally we have been concentrating only on cases where judgements about disagreement are strongest, such as the lack of disagreement between chess opponents and the existence of disagreement between Buba and Kiki in the original case. Things are less clear in other domains. For example, it's not obvious whether or not people can disagree in taste or in emotion. A consequence of our theory is that judgements about whether or not disagreement can occur in these domains are *normative* judgements. To clarify one's thinking about these harder cases, then, one could try to consider what projects people are involved in and whether or not these projects have standards that require alignment. Crispin Wright (2021) partly does this, and it provides a very nice case for a positive answer with regard to disagreement in taste. We think that turning our attention more explicitly to normative considerations will help to provide a fresh angle on these debates.

6 | CONCLUSION

In this paper we have done two things. First, we argued that disagreement without belief is real. We made a theoretical case for this in section 2 and appealed to folk intuitions in section 3. Second, we argued that our normative theory of disagreement can best accommodate this phenomenon. We examined and refuted a selection of other theories in section 4 before presenting our own theory in section 5. Taken together, these arguments give us good reason to believe that disagreement is indeed normative.

REFERENCES

Baker, D., and J. Woods. 2015. "How Expressivists Can and Should Explain Inconsistency." *Ethics* 125:391–424.

Bex-Priestley, G., and Y. Shemmer. 2017. "A Normative Theory of Disagreement." *Journal of the American Philosophical Association* 3, no. 2:189–208.

Blackburn, S. 1998. *Ruling Passions*. Oxford: Clarendon Press.

Bratman, M. 2014. *Shared Agency: A Planning Theory of Acting Together*. New York: Oxford University Press.

Dreier, J. 2009. "Relativism (and Expressivism) and the Problem of Disagreement." *Philosophical Perspectives* 23, no. 1:79–110.

Ferrari, F. 2015. "Disagreement About Taste and Alethic Suberogation." *Philosophical Quarterly* 66, no. 264:516–35.

Finlay, S. 2014. *Confusion of Tongues: A Theory of Normative Language*. New York: Oxford University Press.

Finlay, S. 2017. "Disagreement Lost and Found." In *Oxford Studies in Metaethics*, vol. 12, edited by R. Shafer-Landau, 187–205. Oxford: Oxford University Press.

Frances, B., and J. Matheson. 2019. "Disagreement." *Stanford Encyclopedia of Philosophy* (Winter 2019 Edition), edited by Edward N. Zalta. https://plato.stanford.edu/archives/win2019/entries/disagreement/

Gibbard, A. 2003. *Thinking How to Live*. Cambridge, Mass.: Harvard University Press.

Gibbard, A. 2012. *Meaning and Normativity*. Oxford: Oxford University Press.

Gilbert, M. 2006. *A Theory of Political Obligation: Membership, Commitment, and the Bonds of Society*. Oxford: Oxford University Press.

Kölbel, M. 2004. "Faultless Disagreement." *Proceedings of the Aristotelian Society* 104:53–73.

MacFarlane, J. 2014. *Assessment Sensitivity: Relative Truth and Its Applications*. Oxford: Oxford University Press.

Marques, T., and M. García-Carpintero. 2014. "Disagreement About Taste: Commonality Presuppositions and Coordination." *Australasian Journal of Philosophy* 92, no. 4:701–23.

Priest, G. 2006. *In Contradiction*. 2nd ed. Oxford: Oxford University Press.

Railton, P. 2012. "That Obscure Object, Desire." *Proceedings and Addresses of the American Philosophical Association* 86, no. 2:22–46.

Railton, P. 2018. "Desire, Value, and Learning." John Locke Lectures, University of Oxford. www.philosophy.ox.ac.uk/john-locke-lectures#collapse476186

Ridge, M. 2013. "Disagreement." *Philosophy and Phenomenological Research* 86:41–63.

Ridge, M. 2014. *Impassioned Belief*. Oxford: Oxford University Press.

Searle, J. 1990. "Collective Intentions and Actions." In *Intentions in Communication*, edited by P. Cohen, J. Morgan, and M. Pollack, 401–15. Cambridge, Mass: MIT Press.

Stevenson, C. L. 1944. *Ethics and Language*. New Haven: Yale University Press.

Stevenson, C. L. 1963. *Facts and Values: Studies in Ethical Analysis*. New Haven: Yale University Press.

Sundell, T. 2011. "Disagreements About Taste." *Philosophical Studies* 155, no. 2:267–88.

Sunstein, C. R. 2007. "Incompletely Theorized Agreements in Constitutional Law." *Social Research: An International Quarterly* 74, no. 1:1–24.

Tolkien, J. R. R. 1977. *The Silmarillion*. Edited by C. Tolkien. London: George Allen and Unwin.

Tuomela, R. 2007. *The Philosophy of Sociality: The Shared Point of View*. New York: Oxford University Press.

Tuomela, R., and K. Miller. 1988. "We-Intentions." *Philosophical Studies* 53:367–89.

Wedgwood, R. 2007. *The Nature of Normativity*. Oxford: Clarendon.

Worsnip, A. 2019. "Disagreement as Interpersonal Incoherence." *Res Philosophica* 96, no. 2:245–68.

Wright, C. 2021. "Alethic Pluralism, Deflationism, and Faultless Disagreement." Included in this collection.