White Rose Research Online

eprints@whiterose.ac.uk
https://eprints.whiterose.ac.uk
Universities of Leeds, Sheffield and York

Deposited via The University of Leeds.

White Rose Research Online URL for this paper:
https://eprints.whiterose.ac.uk/id/eprint/195737/

Version: Accepted Version

Article:
Bex-Priestley, G and Shemmer, Y (2017) A Normative Theory of Disagreement. Journal of the American Philosophical Association, 3 (2). pp. 189-208. ISSN: 2053-4477

https://doi.org/10.1017/apa.2017.17

This version is published under a Creative Commons CC-BY-NC-ND licence (https://creativecommons.org/licenses/by-nc-nd/4.0/). No commercial re-distribution or re-use allowed. Derivative works cannot be distributed.

Reuse
Items deposited in White Rose Research Online are protected by copyright, with all rights reserved unless indicated otherwise. They may be downloaded and/or printed for private study, or other acts as permitted by national copyright laws. The publisher or other rights holders may allow further reproduction and re-use of the full text version. This is indicated by the licence information on the White Rose Research Online record for the item.

Takedown
If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

A Normative Theory of Disagreement[1]

Graham Bex-Priestley and Yonatan Shemmer[2]

ABSTRACT: *Expressivists have trouble accounting for disagreement. If ethical or other normative judgments are desire-like rather than belief-like, it is puzzling why we think people often disagree in those domains. While previous expressivists have proposed only straightforwardly descriptive conditions under which disagreement occurs, we argue that disagreement itself should be understood normatively: two or more people disagree just in case their diverging attitudes imply, given a common project of theirs, that at least one of them has reason to change their mind.*

KEYWORDS: metaethics, disagreement, expressivism, non-cognitivism

For realists about a given domain, the phenomenon of disagreement seems easy to account for – someone asserting *p* disagrees with someone asserting *q* if and only if *p* and *q* cannot be simultaneously true (in the sense of accurately representing the world). Since expressivists believe some declarative sentences express non-representational attitudes like approvals or plans, they require a more nuanced account of disagreement.

We present a theory of disagreement according to which ascriptions of disagreement are partly normative. Existing expressivist theories of disagreement such as those offered by Stevenson, Blackburn, Gibbard and Ridge, give only *descriptive* conditions which, they maintain, are sufficient to identify disagreement. On the account we argue for, the claim that people disagree involves the normative claim that at least one party has reason to reconsider their view in light of their common project. If normative anti-realism is true and our arguments are sound, it means there

[1] We would like to thank Luca Barlassina, Lewis Brooks, Alex Duval, David Enoch, Paul Faulkner, Jimmy Lenman, Hallvard Lillehammer, Mike Martin, Veronique Munoz-Darde, Joseph Raz, Daniel Viehoff and our two anonymous reviewers, as well as audiences at the annual conference of the Israel Philosophical Association, Sheffield University and the workshop series 'The Practical, the Political and the Ethical' hosted by the Institute of Philosophy, for challenging comments and helpful advice.

[2] Authors are listed in ascending order of degree of belief in the central thesis (with apologies to Clark and Chalmers).

is no robust fact of the matter about whether two people disagree with one another. We believe our theory offers the best expressivist account of disagreement. However, in this paper we merely aim to highlight its advantages and to convince the reader to consider it as an interesting new candidate for an expressivist theory of disagreement.[3]

1. Desiderata for a Theory of Disagreement

Fundamental normative disagreements are normative disagreements that persist despite agreement on all the relevant non-normative facts. A theory of disagreement that tries to capture the phenomenon of fundamental normative disagreement ought to satisfy the following desiderata:

- It describes the phenomenon of normative disagreement in such a way that justifies our classifying it as a species of the genus 'disagreement'. The essential features that constitute normative disagreement must be shared by cases of theoretical disagreement. Call this feature *'unity'*.
- It matches our pre-theoretical intuitions. That is, it must classify cases we pre-theoretically think of as cases of disagreement as cases of disagreement, and cases we pre-theoretically think of as cases in which no disagreement occurs as cases of non-disagreement. Call this feature *'match'*.
- It explains the possibility of the persistence of normative disagreement despite agreement on all non-normative facts. Call this feature *'persistence'*.

Cognitivists typically explain normative disagreement as a particular case of descriptive disagreement. Persons who normatively disagree, cognitivists assert, disagree in virtue of having beliefs with incompatible contents: they disagree because they describe the world (of reasons) in incompatible ways.

Expressivists cannot subscribe to this cognitivist analysis because, on their view, normative judgments are not descriptive. Instead, expressivists appeal to the notion of conflict between non-cognitive attitudes such as desires, intentions or plans. Their accounts are thus well placed to satisfy *persistence*, since conflicts between non-cognitive attitudes are taken to be independent of agreements in beliefs. On the other hand, they thereby expose themselves to the charge of failing to meet the requirement for *unity* because the concept of conflict upon which they rely is different from the concept of conflict employed in the analysis of descriptive disagreement. Independently expressivists are often accused of failing the requirement for *match*.

[3] We also think that our theory is an interesting candidate for a cognitivist theory of disagreement. Arguing for that claim is not part of our current task.

Before we critically assess existing expressivist accounts, here is a brief overview of our own theory. Our account aims to cover disagreement both in attitudes that play a role in descriptive judgments and in attitudes that play a role in normative judgments. The account is grounded in two basic thoughts. First, that there is no robust fact of the matter about whether or not two given people disagree, and second, that disagreement entails reasons for one of the parties to change their mind. It is designed to satisfy the three desiderata mentioned above. Our main claim is this: *A and B disagree if and only if they have different attitudes and, given a shared project of theirs, this implies at least one of them has reason to change his/her attitude.*

In the following three sections we argue that accounts given by Charles Stevenson, Allan Gibbard and Michael Ridge are problematic. We then present our normative theory of disagreement as an interesting alternative which overcomes the problems of the existing accounts.

2. Disagreement in Preference

Stevenson's (1944) theory of disagreement has been influential for contemporary expressivists. Imagine two people discussing where to go for dinner together. One person is for Chinese food, but the other wants Indian. This clash is viewed by Stevenson as a disagreement about where to dine. Since we may suppose the disputants agree about what food each restaurant serves, the prices, the efficiency of service and all other relevant properties, Stevenson claims the disagreement 'springs more from divergent preferences than from divergent beliefs, and will end when they both wish to go the same place' (Stevenson 1944: 3). On a Stevensonian account, there are therefore two ways for people to disagree. Conflicting beliefs cannot be simultaneously true; conflicting desires cannot be simultaneously satisfied. Blackburn subscribes to a Stevensonian account of disagreement:

> If I am minded to permit smoking in our house, and my wife is minded to forbid it, we do disagree. Only one of these practical attitudes can be implemented, and I am for one, and she is for the other. (Blackburn 1998: 69)

Entering the moral domain, suppose David and Ed are discussing taking military action in Syria. David says it is right that the UK should do so. Ed says it is wrong. On this view, David and Ed are experiencing a clash of attitudes. David is *for* military action in Syria; Ed is against it. Without a change of heart, one of them is going to be unsatisfied with the way the world goes.

Nevertheless, it is implausible to think that all desires with incompatible contents are disagreements. Consider two badminton opponents who both wish to

win the game. Mere wishings are not full-fledged judgments, but we can stipulate that the desire-like state is stronger than a fleeting fancy. Suppose each player has an all-things-considered desire and a plan to win. Both plans cannot succeed; both players cannot be satisfied. It would be odd to label this as a disagreement. In a sense, they both *agree* – they agree that winning is valuable, that it is to be aimed at and strived for, that it is *to be achieved*.

Stevenson seemed aware, at least implicitly, of the problem that mutually incompatible desires are insufficient for disagreement. He sometimes added the extra condition that people only disagree when 'at least one of them has a motive for altering or calling into question the attitude of the other' (Stevenson 1944: 3). This move has parallels with the theory we propose in §5. We believe people with divergent attitudes only disagree when at least one of them has a *reason* for altering or calling into question *their own* attitude (in light of their common project). But Stevenson's own move is far too strong. Atheist Annie disagrees with Catholic Caroline whether she wants to change Catherine's theistic mind or not. And if Annie judges she herself ought to smoke while Catherine thinks Annie really ought to quit, they disagree even if Catherine rather dastardly wishes Annie to retain her poor judgment.

Stevenson's theory of disagreement in preference, then, has counterintuitive consequences. Without the "motive" condition, it wrongly classes badminton opponents as in disagreement with one another, and with it, it rules out the possibility that people can be indifferent or quite happy to remain in disagreement. It fails *match*.

3. Disagreement in Plan

Gibbard believes normative judgments are planning states. To think one ought to turn left at the junction, all things considered, is to plan to turn left if one finds oneself at the junction. Gibbard's theory of normative disagreement is therefore couched in terms of these planning states. A first pass at defining normative disagreement is as follows: A and B disagree in plan if and only if A plans to φ in circumstances C, B plans to Ψ in C, and φ and Ψ are incompatible (Gibbard 2003: 68-71). This overcomes the worry about the badminton players because, were one player to find herself in the circumstances of her opponent on the other side of the court, she too would plan to win. Both players agree that winning is the thing to (try to) do in either circumstance.

But it may seem as though Gibbard will suffer from an inverted case of the badminton match. Suppose you now play a game against one of the original players, but you simply don't care about winning. You're not very competitive and would

actually prefer your opponent to win. You and your opponent seemingly have incompatible contingency plans for your situation – she plans to win if she were in your shoes, and you plan to lose in those exact circumstances. To classify you and your opponent as disagreeing with one another in virtue of planning different actions would be just as odd.

The debate here gets complicated. Gibbard will insist that in this last example you and your opponent are not planning for the same circumstances because you have different goals. If you were in your opponent's shoes *and* her head, you'd plan to win, so you don't disagree after all. We would reply that if the specification of the circumstances must include all of the agent's mental states, it becomes unclear that anyone *could* plan to do anything different than what anyone else plans. Gibbard might suggest a way to include only some of the other person's mental states in the circumstances. We doubt there is a method for deciding which mental states should be included in the specification. We do not however take this objection to be conclusive.

A much deeper problem with Gibbard's account is the strong intuition that divergent plans simply do not count as disagreement. Suppose pro-life Linda plans to never have an abortion whereas pro-choice Isha doesn't rule it out. Both could carry out their plans unhindered by the other, so in what sense are they incompatible? As Gibbard acknowledges, this seems not so much a disagreement but 'a difference of personal characteristics, like having different hair colors' (Gibbard 2003: 270). Why should we care about other people's plans if they don't get in the way of ours?

Gibbard's answer is that we should care what other people think 'because we share thoughts' (Gibbard 2003: 70). We are social creatures. Much of our success as a species is owed to our incredible facility for communication, and this facility has allowed us to let others do our thinking from time to time. Chess players share tactics and strategy without needing to (always) double check each other's reasoning. We are often pressingly in need of "offloading" our reasoning to others, because we are all players in a much more complicated game – the game of living. And it is more complex for two reasons. First, there are more contingencies to plan for. Second, the aim of the game is not declared by fiat.

In thinking how to live we share our life strategies with one another, we learn from our community and we get better at living well. But there is a worry that this runs the risk of reducing our normative differences to non-normative ones. When we share a goal with our neighbour but disagree on the best means to achieve it, it seems as though our disagreement simply boils down to matters of prosaic fact, and it's clear enough why we share thoughts about those. We want the truth, and our neighbours are often just as good as or better than we are at finding it. The deeper question is why

we should take our neighbour's *divergent-at-the-roots* attitude as a disagreement about how to live when, ultimately, we have different goals in mind.

In short, Gibbard's reply is to sing the praises of *treating* conflicting plans as disagreements. We presume he means to treat these conflicts as we would treat disagreements in belief, perhaps by losing some confidence in them. It certainly is true that in most cases we benefit from doubting our plans whenever we learn that someone, whose planning we admire, has made different plans for the same circumstances. If we're planning to take the high road and then notice our esteemed friend is confidently taking the low road, we would ask ourselves "What does she know that we don't?" We would question our plan just like we would question one of our beliefs if somebody, whom we take to be a good authority on the matter, disagreed with us.

But notice the qualification that we respect the judgment of the person we purportedly disagree with. (And this may include thinking her goals are broadly in line with our own.) If we find out upon closer inspection that the person taking the low road is not our esteemed friend but Amy the amoralist who always makes abhorrent decisions, we would *not* question our plan. And Amy, who holds our past decisions in low regard, would not question her plan upon learning about ours. Neither she nor we would see any benefit whatsoever in treating our respective plans as disagreements. She can take the low road if she wants; we're sticking with the high road.

Although he takes pains to stress just how often there is a point to respecting the views of people whose plans diverge from our own, Gibbard acknowledges the possibility of it being truly pointless to engage with someone else's point of view. In these situations, we come to an impasse. And, 'with such an impasse, with no ways left to work toward accord, we could find no point to treating questions of how to live as topics for agreement and disagreement' (Gibbard 2003: 283). We find this far too big a bullet to bite. Whether at an impasse or not, we often think that Linda and Isha disagree about the permissibility of abortion, and this is true no matter how closed minded they are. Indeed, their disagreement may appear deeper as they become more entrenched in their own views. The fact that Gibbard renders this as no disagreement at all is a strong reason not to adopt his account.[4]

[4] See also the discussion in section 5.2 and footnote 16.

4. Disagreement in Prescription

Ridge rejects Stevensonian and Gibbardian accounts of disagreement in favour of basing disagreement in advice. Suppose Kantian Ant thinks a particular killing is wrong whereas Millian Bill judges it to be right. Ridge proposes that we should ask what advice parties would give about how to act in particular circumstances, on the assumption that the advice is given in certain idealized conditions. Since Ant would advise Sue the stabber to not perform the killing whereas Bill would give the opposite advice, Ant and Bill disagree. The idealization involves all parties answering in honesty, full candour and non-hypocrisy. Here is Ridge:

> Two people (or two stages of the same person at different times), A and B, disagree in prescription about D's φ -ing in C just in case in circumstances of honesty, full candour, and non-hypocrisy, A would advise F-ing in C and B would advise Ψing in C, where φ -ing and Ψing are incompatible, in the sense of being impossible to combine without D's thereby being irrational.[5] (Ridge 2013: 60)

Ridge holds that normative disagreement consists in it being 'impossible to follow both pieces of advice' (Ridge 2013: 61). One problem with this account is that people may not give advice in accordance with their moral judgments.[6] Suppose Bill, who thinks it right for Sue to perform the killing, judges that *advising killing* would diminish utility and that advising not-killing would maximise utility. Perhaps he thinks that society will see him as a monster if he advises killing, for instance, or perhaps he thinks Sue will do the opposite of what he advises. Even Ant may not advise Sue not to murder. As a Kantian, perhaps Ant sees more value in letting Sue come to the decision autonomously rather than under the authority of a friend's prescription. After all, it is quite common for people to think it wrong to advise an action in certain circumstances, even if the advice is to do what they judge to be right.

Ridge may respond that these are cases of dishonesty or lack of candour. Bill (with good reason) is not being honest when he fails to advise Sue to commit the murder and Ant is not being fully candid when he fails to advise her not to murder. You might wonder what honesty and candour amount to. Although one cannot make sense of "lying" when it comes to prescriptions due to their not being truth-apt, one

[5] Ridge adds the final clause about becoming irrational in order to account for disagreement in belief. A and B might disagree about D's *believing that p* in C. On Ridge's account, this could happen because A would advise D to believe *p*, B would advise D to believe not *p*, and while it may be possible for D to follow both pieces of advice and believe both *p* and not *p*, D cannot do so without thereby being irrational. In this way, Ridge satisfies our desideratum of *unity*.

[6] Eriksson (2015) has a similar worry about advice diverging from judgment, although his worry is specific to Ridge's particular theory of the (hybrid) mental states that constitute normative judgments.

may talk of sincerity. Let us stipulate that advice is sincere (both honest and candid) if and only if the speaker advises what they want their audience to do.

Unfortunately, this still renders Ridge's account problematic. By hypothesis, Ant wants Sue to not murder of her own accord. It seems that the only kind of advice Ant can sincerely give Sue is to tell her to think through the issue more carefully. Ridge may invoke the notion of "non-hypocrisy" at this point, but this is troubling for a different reason. If being non-hypocritical means one's advice must be to do exactly what the speaker would do in that situation, then it seems as though the account collapses into Gibbard's. For suppose Ridge says "Imagine we gave Ant a truth (or sincerity) serum and asked him a straight question." The problem is this: what question are we asking Ant? If it is "What advice would you give Sue if she was given the opportunity to commit the murder?" then Ant would answer "None. I certainly wouldn't tell her to avoid murdering because the decision has to come from her." Instead, Ridge seems to be asking what advice Ant would give if he *had* to give some specific advice, and if Sue *had* to follow it, and if nobody – not Sue, not the community, not even himself – would ever find out what advice he gave. This is starting to sound very much like asking Ant what he would make Sue do if he could control her like a puppet, which in turn sounds very much like a personal planning question: what would *you* do if you were Sue? If that's what Ridge's notion of advice boils down to, the difference between his account and Gibbard's first pass is very unclear.

Ridge therefore faces a dilemma. He can leave us with a fairly loose understanding of what would be advised in the closest worlds where we give sincere advice to an agent, but this fails to satisfy *match* - Ant and Bill's advice for Sue may not conflict, despite their disagreement on the moral status of her murdering. Alternatively, Ridge can tighten the conditions such that we are forced to advise "Do exactly as we would do." While this satisfies *match*, it suffers from the same problem Gibbard identified with his first pass: the mere fact that we would do different things in a given situation looks more like a difference than a disagreement. This is what motivated Gibbard's story about why it is almost always beneficial for us to *treat* our normative differences as disagreements, which led to his failure to account for cases of impasse. Could Ridge give a different story?

Ridge proposes that two people with diverging plans for the same circumstances disagree because if they were to give sincere and non-hypocritical advice to a person, the advice that each one of them would give in these circumstances could not be followed at the same time. But why is that supposed to make us think that they are involved in disagreement? Gibbard's story was that if two people were to have an interest in learning from each other about how to live their lives then their diverging plans would constitute disagreement, since at least one of them would have a reason to reconsider their plan given their respect for the other as a good planner. The problem with Gibbard's answer was that it applied only contingently. In

circumstances of impasse people would rationally lose the desire to learn from each other's plans. Ridge's answer depends in the same way on a contingent common project. If two people share the project of sincerely advising a single person then they have reason to find a way to offer non-diverging advice. But there is nothing necessary about the project of giving advice to the same person.

4.1 Where do things stand?

Stevenson, Gibbard and Ridge's theories all satisfy *persistence*, since according to all three it is possible that disagreement will persist beyond agreement on all non-normative facts. Stevenson's theory fails *match.* The extension of what the theory counts as disagreement does not match our intuitions. As Ridge points out Stevenson's theory fails *unity* too (Ridge 2013: 46). Gibbard's account also fails *unity* and *match*. His theory is not applicable to disagreement in belief and he fails *match* when it comes to cases of impasse between those who fundamentally disagree.

Finally, Ridge's account satisfies *unity*, but is then impaled on the horns of a dilemma. On the first horn he allows advice to diverge from judgment and he fails *match*. On the second horn he amends the theory so that the advisor must advise what he himself would do in the situation. This satisfies *match* but it leaves us with the same puzzle that Gibbard faced initially, namely that a mere difference in what someone would do does not constitute disagreement.

5. A Normative Theory of Disagreement

We have seen how previous theories of disagreement given by expressivists have failed to match our intuitions. Either they are too demanding and we can imagine cases of disagreement where the suggested necessary conditions are not satisfied, or they are too permissive, construing situations as disagreements when they appear not to be. The attempt to capture judgments of disagreement as descriptions of a relation between attitudes seems to be missing something crucial.

What's missing? Gibbard's discussion of impasse gives us an important clue. In his attempt to convince the reader that cases of apparent impasse are cases of real disagreement, Gibbard aims to show that even in these cases each party has reason to take seriously the views of the other. He aims to show that, even in situations where the parties to the apparent conflict share no common fundamental values, they *should* see the clash in attitudes as undermining the justification for their own view. Gibbard is unsuccessful, but the attempt itself is revealing: he is implicitly relying on the idea that disagreement (at least if we have no independent reason to think the other party is uninformed, irrational, or otherwise epistemically inferior) gives us *reasons* to

reconsider our own judgments.[7] The view is also familiar from realist discussions of normative disagreement and more generally from discussions of non-normative disagreement.[8] The realist, however, has an easy explanation of why disagreement indicates that at least one person should change their mind: if parties make conflicting judgments, at least one person is incorrectly representing the world and therefore has reason to alter their view (with a proviso to be discussed shortly).

On the other hand, expressivist accounts of disagreement fail because they cannot explain the rationale for such normative pressure to adjust one's own judgments in light of disagreement. The mere fact of a clash between non-cognitive attitudes cannot rationalize this normative pressure, as we argued in earlier sections of this paper. After all, on the expressivist view fundamental non-cognitive attitudes are not meant to respond to or 'fit' with any normative facts and so disagreement cannot be an indication of a failure of fit.

It is instructive to consider the realist's rationale for the demand to reconsider one's normative judgments in cases of disagreement with epistemic peers. The realist's intuition that disagreement with an epistemic peer is grounds for reconsideration of our own judgments is itself the manifestation of an implicit assumption – the proviso mentioned earlier. Realists implicitly assume that the disagreeing parties share a common project, namely, the common search for normative truths.[9] It is only because the other party to the disagreement shares this goal, and is therefore guided by the same standards, that their view ought to be taken as a reason for adjusting one's own.[10] The idea of tying disagreement to a shared purpose is hinted at by Stephen Finlay, who thinks the 'relevant factor [for disagreement] is evidently the *conversational ends* of the speakers' (Finlay 2014: 131). Attitudes of one party, in domains in which there is no shared goal, regardless of what the relation between them is – regardless, for instance, of whether they can rationally coexist in the mind of a single agent – cannot put pressure on the other party to adjust their own.[11] Thus judgments about disagreement include, even on the realist analysis,

---

[7] We have considered here the argument Gibbard gives in *Thinking How to Live* (particularly 2003: 280-283). He offers a different argument for the same conclusion in *Wise Choices, Apt Feelings* (1990: 180). See also Shemmer (2014) for a criticism of that argument.

[8] Folley (2001: chapter 4).

[9] The common project may not be optional: according to some authors aiming for truth is constitutive of beliefs (Velleman 2000: introduction).

[10] Imagine (if you can) a being whose "beliefs" did not aim at the truth but something else – humour, say. This creature "believes" there is a purple tortoise in the room practising ballet; you do not. The fact that your attitudes do not line up gives neither of you a reason to alter them, and this is because you have different goals with different standards for success.

[11] Of course, we might have instrumental reasons to adjust our attitudes in light of those of others even when we do not share a common project, e.g. if we learn of a clash of attitudes and are afraid of being hurt if we don't adjust our own, but our focus here is on the claim that the divergence implies at least one of us has reason to change our attitude *merely* in light of our common project.

an implicit judgment about a shared project, and therefore about standards that ought to guide the disputing parties. This gives us the necessary anchor for our new account of disagreement.

We propose that judgments of disagreement are in part normative judgments – normative judgments about the standards that govern the formation and retention of attitudes of the disputing parties and therefore about the reasons that apply to the disputing parties. Since on the expressivist analysis normative judgments are expressions of non-cognitive attitudes, judgments of disagreement are also best understood as, in part, expressions of non-cognitive attitudes held by the person who makes the judgment that certain parties disagree.

The view suggested here departs in this respect from all former expressivist analyses of disagreement. On all former expressivist analyses of disagreement, judgments that people are in disagreement are *descriptive* judgments. They describe a certain relation between the non-cognitive attitudes of the disputing parties. This descriptivist approach is at the heart of the failure of expressivist accounts of judgments of disagreement. As we have seen, no set of purely non-normative conditions in the offing has been necessary and sufficient for disagreement. Expressivists must stop seeing judgments of disagreement as purely descriptive and recognize their normative component.

We aim to offer a unified analysis of disagreement which will cover disagreement in normative and descriptive judgments, and possibly even disagreement in attitudes other than the ones that play a role in descriptive and normative judgments.

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.[12]
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.[13]

To illustrate, suppose Alice and Bob are about to paint a fence. Alice plans to paint it green all over but Bob plans to paint it blue. Due to this divergence in attitude combined with their common project of painting the fence, there is reason for either Alice or Bob (or both) to change plans. If they don't, they won't get very far. This is not to say that either Alice or Bob has *most* reason to change plan. Perhaps they are

[12] This may or may not be a normative judgment. We take no stand on this issue.

[13] The analysis can be naturally extended beyond two parties judged to be in disagreement.

both being paid to paint the fence their respective colours, or threatened if they fail to do so. However, these are extraneous considerations. The conditions are met, so Alice and Bob disagree – they disagree about what colour to paint the fence.

5.1 Clarifications

Cognitivists might also adopt our analysis, but for now we focus on the advantages it offers expressivist. Due to the normative component in our analysis, and given the expressivist understanding of normative judgments, an expressivist who adopts it will maintain that there are no robust facts about whether two people disagree. Judgments of disagreement, she will claim, are in part expressions of desire-like attitudes.

One might have the intuition that even if expressivism is true there must be robust facts about disagreement. For instance, if Geoff thinks grass is green and Jingbo thinks grass is green, it is a robust fact that they do not disagree.[14] However, at least in this particular case, we can agree that it is a robust fact that Geoff and Jingbo don't disagree because the fact they don't disagree does not depend on what reasons they do or don't have. Condition #1 is a simple description, and one that is robustly not met by Geoff and Jingbo. Because disagreement contains both descriptive and normative components – disagreement is *thick* – there can indeed be robust facts about when there is *no* disagreement. On the other hand, in cases where the descriptive conditions are met, our theory does imply that expressivists are committed to there being no robust facts about whether there is or isn't disagreement.

Our analysis is not tied to any particular expressivist account of normative judgment (or any account of normative judgment at all). If our analysis of disagreement is correct and if expressivists are right about normative judgments then judgments of disagreements involve the expression of *some* desire-like state. We remain neutral on exactly what state that will be.

The reason mentioned in condition #3 must be grounded in the project A and B are undertaking, and given this project, the existence of the reason must be implied by the difference in their attitudes. An example will help. Suppose A believes London is the capital of Kenya and B believes Paris is the capital of Spain. They have a common project of finding the truth, and at least one of them (in fact both) has *pro tanto* reason to change their belief (because it's false). But the reasons they have to change their attitudes are not implied by the difference between them, so this does not count as a disagreement. On the other hand, if B believes London is not the capital of Kenya, *this* implies that either A or B has a false belief and so at least one of them has reason to change their mind – A and B would disagree about whether London is Kenya's capital.

[14] Thanks to Paul Faulkner for this point.

To know that A and B disagree, we needn't know *which* of A and B has reason to change their attitude. Furthermore, A and B need not think of themselves as in disagreement to actually be in disagreement. They may not even know about each other's existence. This also explains why the normative component of our analysis is given in terms of reasons and not rationality, since disagreement is possible between parties who are not aware of each other and who have no access to the reasons they might have to change their mind. Reasons are to be understood objectively here. All A's available evidence might support A's belief that *p* and all B's evidence might support B's belief that not *p* so neither of them has subjective reason to think otherwise. One might object that our analysis of disagreement entails an over-intellectualized understanding of judgments of disagreement. That is not so. To judge that people disagree need not involve any explicit judgments that our three conditions hold. The three component parts could be tacit judgments, and often the second and third may well be "default" judgments – that is, the agent need not process them even sub-agentially, unless some defeaters make her doubt that the parties share a project or that the standards set by the project require alignment of attitude.[15]

Finally, what is a common project? The question has two parts. What is a project? And what is a common project? A 'project' is another word for a plan or a goal. There are multiple answers in the literature to the question what it is to have a goal. Davidson would have identified one's goals with the propositional objects of one's intentions. Some authors identify one's goals with the propositional objects of the desires one *would* have were one consistent, whereas for others one's goals are identified with the propositional objects of those of one's desires that are endorsed by second-order desires (Frankfurt 1969) or coherent second order self-governing policies (Bratman 2000). Furthermore, some projects may not be optional. Perhaps our beliefs always aim at truth no matter what. Perhaps the project of living together ethically is non-optional, and so we are always bound by the standards of morality regardless of our particular desires. We stay neutral on the question of what determines one's projects. A common project is a project that two or more people have in common. This does not imply the agents are acting jointly, but merely that they are bound by the same standards. We are neutral on the question of what these standards are and whether they are normatively determined, constitutively determined, or something else.

5.2 The advantages of a normative theory of disagreement

Our analysis satisfies the three main desiderata of a theory of disagreement: *unity*, *persistence* and *match*. Let us first see how our theory allows us to give a unified account

[15] Thanks to Joseph Raz and Alex Duval for pushing us to clarify this point, and thanks to Luca Barlassina for the suggestion to appeal to default judgments.

of normative and descriptive disagreement. On our view, both normative and descriptive disagreement is partly constituted by normative truths. When A and B descriptively disagree, they have different attitudes (beliefs), they are engaged in a common project (the project of finding out the truth), and given the standards governing this project, their differing attitudes imply at least one of them has a reason to change it (because it implies at least one of their beliefs is false). As the same analysis works for normative disagreement, our account is unified.

Furthermore, it has the capacity to extend to other attitudes. Someone may sincerely consider two idealists who differ in their *hopes* to be in disagreement, for instance. If we think she is wrong, our disagreement with her may be normative – we may think the standards governing appropriate formation and retention of hopes do not require them to be in alignment. To answer Ridge's charge, our account explains what unifies the states of mind that are apt for disagreement: they are those we may have reason to alter in light of divergence and a common project.

The account also satisfies the desideratum of *persistence* for disagreement in the normative domain if expressivism is true. Two instrumentally rational people may disagree about normative matters, despite their agreement on all non-normative facts. This is because, first of all, they may have different desire-like attitudes despite having all the same representational beliefs. Secondly, this difference under a common project may imply at least one of them has reason to change their attitude. This is, of course, a normative judgment, and one that others may dispute. The debate between reasons internalists and reasons externalists may strongly bear upon our intuitions about normative disagreement.

Indeed it is an advantage of our account that it predicts that if parties in a purported normative dispute believe the normative domain is a domain of robust facts, then they will themselves view the disagreement as implying a reason for at least one of them to reconsider their attitude; whereas if they do *not* see the normative domain as a domain of facts, they may, but need not, think of their differing attitudes as providing neither of them with a reason to reconsider their attitudes. Of course a third party might at the same time think that it does imply such a reason.[16] On our view, the battle between those who view divergence in conative states as potential disagreements and those who do not is to be fought on normative terrain.

Lastly, our account satisfies *match*. Recall Stevenson's minimal theory where disagreement was a matter of parties having beliefs, or desires, with inconsistent contents. This counter-intuitively classed badminton opponents as in disagreement, since both of their desires to win could not be satisfied. Our account explains why

[16] An instance of this can be seen in the debate about whether cases of fundamental impasse imply at least one person has reason to change their view. While we are tempted to agree with Ridge (2013: 53-54) that it does, Gibbard (2003: 283) thinks it does not.

badminton opponents do not disagree – they have no reason to alter their desires for victory in light of their common project of playing badminton. Badminton is a case of a "good" clash of attitudes, whereas preferring different restaurants (under the project of dining together) is a "bad" clash. And in contrast to Stevenson's extra condition that at least one party wishes to change the other's mind, our account allows that A and B disagree even when neither A nor B desires the other to change their attitude. As opposed to Ridge the account allows us to divorce disagreement from advice. A and B can disagree in their diverging attitudes even while giving the same advice, or agree while giving differing advice. Finally, as we have seen in our discussion of *persistence* in the previous paragraph, our account allows us to claim, as opposed to Gibbard, that in circumstances of impasse a third party may judge that at least one of the two disputing parties has reason to reconsider their attitudes, even if the disputing parties themselves see their case as one of mere difference and therefore as one in which neither party has reason to reconsider their attitudes.

6. Objections and Replies

In this section we respond to four potential objections to our normative theory of disagreement.

6.1 Unreasonable projects

A and B are both white supremacists and have different views about the best way to promote their ideology. However, their ideology is misguided and is not a source of reasons. White supremacy gives A and B no reasons whatsoever, and therefore no reason to change their minds given the difference in their views in light of this project. It seems that our view thus entails that A and B are not in disagreement. But clearly, says the objector, they do disagree.

An initial reply is that A and B disagree in belief since their different views give at least one of them a reason to reconsider given their common project of finding the most efficient way to promote white supremacy. Of course the disagreement may be expressed in terms of what to do, but their judgments about what to do rest on different beliefs (about how best to achieve their goals). The disagreement about what to do is derivative – A and B have no disagreement about fundamental normative judgments.

To get to the heart of the matter the objector may specify the case as follows: A and B are both white supremacists but disagree about whether it is best - not instrumentally but rather morally - to first exterminate race R1 or to first exterminate race R2. Since their racism is misguided the standards of their racist project give A and B no reasons whatsoever and therefore no reasons to change their minds, given the difference between their views in light of this unreasonable project. So, the objector

will argue, according to our analysis A and B do not disagree. But it certainly looks like they disagree.

If the divergence between the views of A and B is not the result of a disagreement about how best to bring about a certain result, namely the disappearance of all non-white races, then the difference in their position is a difference in fundamental norms. A supports white supremacy of type i and B of type ii. Given this understanding, the project they share is the general moral project and they have different views about what fundamental principles should guide this project. As long as there are general moral standards that apply to A and B and since given these standards the difference in views of A and B is an indication that at least one of them has reason to reconsider their position, our analysis entails that they do disagree. One way to believe that the difference in views of A and B is an indication that at least one of them has a reason to change her view is to think morality is about getting to the moral truth, and so one of the divergent views must be incorrect and thus ought to be altered to represent reality. A different way is to think morality *just is* the business of aligning goals and finding ways to live together, and so a divergence in moral opinion implies at least one person should change in order to live in harmony.

What if A and B share the project of morality but it turns out there are no moral standards; that the project of morality is pointless and therefore not a source of reasons? The objector may claim that in this case our analysis entails that A and B are not in disagreement even though they still seem to be in disagreement. In response, we would reply that even if the project of morality is pointless that does not mean that there are no common standards imposed on A and B. As long as A and B have both committed to a common plan (the plan, say, of finding ways of living together) these commitments may impose standards on them which, given their diverging views, imply at least one of them has reason to change his view.[17]

The objector could however insist that we consider a world in which no standards of any sort apply to A and B in light of which, given their diverging views, they have reason to reconsider. In that case there are no reasons that apply to A and B, neither reasons of morality nor reasons of any other sort, and therefore no reason not to diverge in attitudes. The objector could then argue that while, given our analysis, A and B do not disagree, intuitions still suggest they do.

[17] It may be the case that standards of rationality apply to them in light of which their difference in attitudes indicates that at least one of them has reasons to reconsider. In other words we might think that just as in the *intra*personal case a plan imposes on a person standards of consistency, so in the *inter*personal case a common plan imposes standards of interpersonal consistency. To illustrate, even if there is no good reason to decide to paint the garden fence, you might think that *if* two people decide to paint it, they had better align their opinions on what colour paint to use.

We do not share the objector's intuitions. In a world in which A and B have no reason whatsoever to cohere in attitudes our intuition that A and B disagree vanishes. In such a world the difference in attitudes of A and B will seem like the difference of attitudes of people who like different flavours of ice cream and who insist they face disagreement.

So far our objector stipulated situations in which parties do or do not have reason to change attitudes. But expressivists deny that there are robust facts about practical reasons. On their view claims about practical reasons are expressions of non-cognitive mental states that might vary from one person to another. An expressivist who adopts our analysis of disagreement will predict that judgments of disagreement will vary with the observer's judgments about reasons. This in turn explains why there are conflicting opinions about whether people with conflicting judgments who are at an irresolvable impasse are really in disagreement.

We should note that the term 'disagreement' may have a purely descriptive usage of sorts. A visitor from a different culture may use the term 'disagreement' to refer to disputes that are referred to by a local community as cases of disagreement. The foreign visitor may adopt this terminology in order to simplify communication with the local community. In that case the visitor uses 'disagreement' as shorthand for 'cases that may or may not be disagreement, but you locals see as disagreement'. Alternatively one might use the term 'disagreement' to refer to divergences that would be disagreement if the values held in common by the disputing parties were, counter to the fact, reason-giving. In that case 'disagreement' is used as shorthand for 'cases that in fact are not disagreement, but would be disagreement if the purported values accepted by the disputing parties were in fact reason-giving'. Such elliptical usages should not confuse our understanding of what disagreement is.

Finally, the objector might contend that while an expressivist had better concede that if no norms apply to A and B we would lose the intuition that A and B disagree, that concession does not support an expressivist account of disagreement over non-expressivist accounts. About that we agree with the objector. Our task here is limited. We want to show that our account is a good expressivist account of disagreement and an interesting candidate for a general account of disagreement. We do not intend to show that it is superior to all non-expressivist accounts.

6.2 Conciliating steadfast epistemologists

Sam believes the train is due at five past. Isela, her epistemic peer, believes it is due at ten past. Neither of them has any special reason to think one is in a better position than the other with regards to the truth. In the face of peer disagreement such as this, epistemologists may think it is rational either to be *conciliatory* or to remain *steadfast*. Conciliatory views (Elga 2007, Feldman 2007) propose that it is rational for Sam to

lower her confidence upon learning that Isela disagrees. Steadfast views (Rosen 2001, Kelly 2005) hold it to be rational for Sam to stick to her guns and retain the strength of her belief. On the face of it, it seems as though steadfast views threaten our theory. If neither Sam nor Isela has a reason to reconsider their beliefs in light of their divergence and their common project of finding the truth, our theory classifies them as not in disagreement. But they are. This would mean we fail to satisfy *match*.

Discussion of conciliatory and steadfast views is normally couched in terms of what it is *rational* to believe rather than what we have *reason* to believe. Our answer therefore involves a characterisation of steadfastness that we hope is plausible to its defenders. When translated into the language of reasons, we believe the steadfast view is best characterised as the view that the (mere) fact that an epistemic peer's belief conflicts with your own is not a reason to alter your credence. This is compatible with our view. We are only committed to the conflict *implying* that at least one party *has* a reason to change their mind – not that the conflict *is* the reason.

If the steadfast view is true, what reason is there for either Sam or Isela to change their mind? Reality. If the train is not in fact due at five past, Sam has an objective reason to alter her belief to reflect reality. If the train is not due at ten past, Isela has reason to change her mind. The train is either not due at five past or not due at ten past, therefore at least one of the two parties has a reason to change their view.

Descriptive disagreements come cheaply, then, but this is no objection to our theory. There is very little controversy about when two people disagree in their beliefs. Our theory unifies disagreement in belief with disagreement in other attitudes and it also explains why it is easy to tell when two people with conflicting beliefs disagree – beliefs aim at the truth, there is plausibly only one 'truth' to aim for,[18] and so if two beliefs conflict, at least one of them does not match reality. This implies at least one person has reason to change their mind, assuming the project of finding the truth is reason-giving.

When it comes to non-representational states, the mere existence of any disagreements at all is controversial. This, we believe, is evidence in favour of our theory. The controversy boils down to whether there is reason to align our non-cognitive attitudes such as plans and desires when there is no robust reality we are aiming to represent. This was why Stevenson sometimes claimed that parties must wish to change each other's mind if they are to be classed as in disagreement, and why Gibbard spends a sizeable portion of his book arguing that we have reason to treat each other's plans as learning points for our own. Our insight has been to consider common projects that require alignment of conative attitudes.

To summarise, in cases of disagreement the reason to change at least one attitude is *implied* by the diverging attitudes in light of a common project, but the

[18] Of course, if we think truth is relative, our intuitions about disagreement get murkier, and rightly so.

reason need not *be* the fact that the attitudes diverge. This move allows us to make room for steadfast views in social epistemology.

6.3 Levels of specificity

Suppose A believes murderers should be punished but suspends judgment about how. B believes murderers should be executed. A and B have different attitudes, and it is plausible that they share a common project – perhaps the project of determining just punishment, or the general moral project of finding rules and regulations to live by. Some may have the intuition that the divergence implies that at least one of the two has reason to change their mind. Their inquiry certainly doesn't seem finished. It might be natural, then, to think our three conditions are met. However, intuitively A and B don't disagree because their judgments are clearly compatible.

There are two different ways to interpret this objection. On the first our conditions are not met, whereas on the second it begins to look like A and B do actually disagree. On the first interpretation, A's judgment that murderers should be punished is perfectly compatible with B's judgment that they should be executed. After all, execution is a form of punishment. Therefore our third condition is not met – the mere fact that A and B have these (admittedly different) attitudes does not imply one of them has reason to drop it. A could *retain* her attitude while later adopting B's. B could retain his attitude and he presumably already shares A's.

On the second interpretation, we might take A's suspension of judgment more strongly. It is not that A simply hasn't decided on the specifics yet, but that A *rejects* forming a judgment, perhaps on the basis that the available evidence is insufficient to warrant favouring a specific kind of punishment. In this case, we believe A's attitude – that of thinking murderers should be punished while *rejecting* further specificity – along with B's that the evidence warrants favouring the death penalty does indeed imply that either A herself or B has reason to change their mind. Our three conditions are met. Yet the case has now been described in a way that makes it intuitive to think of it as a case of disagreement: A and B disagree about what the available evidence warrants.

6.4 Judgments of disagreement

We have presented a theory of disagreement and claimed that expressivists would do best to adopt this theory. Since the theory consists in part of a normative condition, an expressivist who accepts it will subsume the analysis of disagreement under her general account of normativity. She will thus, at least in the first instance, speak not of

disagreement itself but rather of *judgments* of disagreement,[19] and will understand these judgments as consisting in part in the expression of non-cognitive attitudes. The current objection targets the resulting account of judgments of disagreement.

Consider the following scenario: a judge J believes that A is for giving money to charity and B is for lowering taxes, and therefore that their attitudes a and b are different. J also judges that A and B share a common project of morality and that (for some reason) 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 attitudes. An expressivist who accepts our analysis of disagreement, says the objection, is bound to claim that in this situation, J would judge that A and B disagree. But this does not seem right.

The objection's apparent success is grounded first in the thought that giving money to charity is not incompatible with lowering taxes, and second in the intentionally vague characterization of J's reasons for thinking that one of A and B has reason to change their attitudes. We reply that once J's judgment about reasons for changing attitudes is clarified the objection dissipates. If for example J believes that the project of morality requires people to have similar attitudes to money and further believes that lowering taxes is bound to make people greedy and to prevent them from giving to charity, then it seems plausible that J would judge that A and B disagree. If on the other hand J is described as a person whose judgment about reasons is utterly unintelligible, then we are unlikely to share his judgment of disagreement and are also incapable of having reliable intuitions about whether or not he would judge disagreement.

Additionally, as we mentioned in section 5.1, an expressivist who accepts our analysis of disagreement and who incorporates such analysis into her general account of judging reasons, is not thereby providing an account of the mechanisms that ground judgments of disagreements. She is not committed to the claim that people explicitly or implicitly apply our three conditions when judging disagreement. Our expressivist is merely committed to the claim that instances of disagreement are best understood in terms of our analysis; an analysis that, for all we have said, is opaque to the person who judges disagreement.

7. Conclusion

Stevenson, Gibbard and Ridge each fail some of the desiderata for a theory of normative disagreement. Gibbard and Ridge's failures are instructive. In order to meet *match* both theories include an implicit appeal to a project common to the disagreeing

[19] To examine the judgment rather than the metaphysics of the thing itself is what Dreier (2015: 273) calls "the expressivist sidestep" and what Carter and Chrisman (2012: 324) call "the core expressivist manoeuvre."

parties. However, both projects are contingent, and the realization that this is so reveals a lacuna in both views: neither can explain why in the absence of either common project diverging conative attitudes constitute disagreement and not difference.

We suggest a theory of disagreement which is in particular a normative theory of disagreement. On our view the judgment that people disagree is itself a normative judgment, in part to the effect that at least one of the disagreeing parties has reason to change her attitude in light of their diverging attitudes and the common standards imposed on her by their common project. Our account meets all three desiderata for a theory of normative disagreement. Seeing judgments of disagreement as normative judgments also allows us to explain the key connection between the common project that disagreeing parties are engaged in and the normative pressure to which they are subject to adjust their attitudes in light of their disagreement. Furthermore, it allows us to provide this explanation without assuming that the disagreeing parties will always consider themselves bound by a common project.

We offer this theory primarily as a substitute to existing expressivist theories of disagreement, but it could equally be adopted by cognitivists. In future work we aim to show that it is of particular interest to subjectivists, since it would allow them to meet *persistence* without giving up the cognitivist nature of their position.

References

Blackburn, S. (1998) *Ruling Passions: A Theory of Practical Reasoning*. Oxford: Clarendon Press.

Bratman, M. (2000) 'Reflection, Planning, and Temporally Extended Agency'. *The Philosophical Review*, 109, 35-61.

Carter J. A. and Chrisman M. (2012) 'Is epistemic expressivism incompatible with inquiry?' *Philosophical Studies*, 159.

Dreier J. (2015) 'Explaining the Quasi-Real'. In Russ Shafer-Landau (ed.), *Oxford Studies in Metaethics*, Volume 10 (Oxford: Oxford University Press), pp. 273-96.

Elga A. (2007) 'Reflection and Disagreement'. *Noûs*, 41(3), 478-502.

Eriksson, J. (2015) 'Explaining Disagreement: a Problem for (Some) Hybrid Expressivists'. *Pacific Philosophical Quarterly*, 96, 39-53.

Feldman R. (2007) 'Reasonable religious disagreements'. In Louise Antony (ed.), *Philosophers Without Gods: Meditations on Atheism and the Secular* (Oxford: Oxford University Press), pp. 194-214.

Finlay, S. (2014) 'The Pragmatics of Normative Disagreement'. In Guy Fletcher and Michael Ridge (eds.) *Having it Both Ways: Hybrid Theories and Modern Metaethics* (New York: Oxford University Press), pp. 124-147.

Foley, R. (2001) *Intellectual Trust in Ourselves and Others*. Cambridge: Cambridge University Press.

Frankfurt, H. (1971) 'Freedom of the Will and the Concept of a Person'. *The Journal of Philosophy*, 68(1), 5-20.

Gibbard, A. (1990) *Wise Choices, Apt Feelings*. Cambridge: Harvard University Press.

- (2003) *Thinking How to Live*. Cambridge: Harvard University Press.

Kelly T. (2005) 'The epistemic significance of disagreement'. In John Hawthorne & Tamar Gendler (eds.) *Oxford Studies in Epistemology*, Volume 1 (Oxford: Oxford University Press), pp. 167-96.

Ridge, M. (2013) 'Disagreement'. *Philosophy and Phenomenological Research*, 86(1), 41-63.

Rosen, G. (2001) 'Nominalism, naturalism, epistemic relativism'. *Philosophical Perspectives*, 15, 69-91.

Shemmer, Y. (2014) 'On the Normative Authority of Others'. *Philosophia*, 42(2), 517-21.

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

Velleman, J. D. (2000) *The Possibility of Practical Reason*. Oxford: Oxford University Press.