Normative Disagreement for Expressivists

Louis Bernard

Submitted in accordance with the requirements for the degree of Master by Research

The University of Leeds
School of Philosophy, Religion and the History of Science

September 2025

Intellectual Property and Publication Statements

I confirm that the work submitted is my own and that appropriate credit has been given where reference has been made to the work of others

Acknowledgements

I am grateful to my supervisors Will Gamester and Jack Woods for their practical guidance, patience, and moral support throughout the last year. My meetings with both Jack and Will always motivated and inspired me.

Abstract

The present thesis argues that expressivists can account for normative disagreement in terms of the inability for normative judgements to jointly bring about mental states in a particular agent. I explore existing expressivist-friendly accounts of normative disagreement and disagreement-in-general, and generate success criteria based on their respective successes and failures. I then argue for a particular theory of normative disagreement, demonstrating that it meets these criteria, and succeeds against a number of possible objections that may be raised against it.

Contents

1. Introduction

What is expressivism? To be an expressivist about moral judgements, for example, is to deny that moral judgements are ordinary descriptive beliefs. Under this view, these judgements don't tell us anything about how the world is. They are sometimes characterised as desire-like attitudes. As such, they are not asserted, as facts, but expressed. Many expressivists about morality will also be expressivists about other types of attitudes. Normative expressivists take any 'ought' or value-judgement to be desire-like in this way. Normative expressivism will be my focus in this essay. In particular, I will try to explain disagreement between attitudes in a way that is friendly to normative expressivism.

The worry that I will address in this essay, simply put, is that there is no adequate explanation of why two normative attitudes disagree that does not appeal to the truth or falsity of these attitudes. So what responses have expressivists offered in response to this objection? Previous explanations of disagreement have been numerous and diverse, in part because there is no unified expressivist theory. This is to say that even among those who agree that some attitudes, like moral ones, are non-representational and non-cognitive, there is no agreement on how exactly this is so. Perhaps the greatest reason for this diversity in expressivist theories is the prevalence of the Frege-Geach problem. Like expressivism itself, the Frege-Geach problem no longer has just one formulation, but one way in which the challenge can be framed, known as the Negation Problem (Dreier, 2006), is as a demand for an explanation of what negation is without appealing to the truth or falsity of the attitude being negated. Expressivist theories that offer different solutions to the Frege-Geach problem might well be committed to different solutions to the problem of explaining disagreement. This is because the problems seem to share a lot of theoretical space - like the need to explain inconsistency. In spite of this, this project does not concern the Frege-Geach problem directly. A variety of expressivist-friendly accounts of disagreement will be reviewed, some of which are embedded in fully-fledged expressivist theories, others are simply compatible with expressivist theories of normative judgement. From the review of existing accounts, I will extract some criteria that a successful account of disagreement needs to meet. These will be based on the respective successes and failures of existing accounts to explain disagreement and to predict disagreement in the right sets of circumstances. Ultimately, this will serve to inform and aid my defense of a particular expressivist account of normative disagreement.

The focus of this essay is normative disagreement. This encompasses moral disagreement, which is of course particularly important to metaethical expressivists. However, it also applies to other value-judgements, like aesthetic ones, and all-things-considered ought judgements. As I will explain in the following section, normative disagreement is not isolated from other types of disagreement, and an account of normative disagreement will be subsumed by a broader account of disagreement in general. As such, there will be some reflective equilibrium when evaluating accounts of disagreement in general, normative judgement, and normative disagreement. An account of normative judgement cannot be too revisionary of normative disagreement, and of disagreement in general. Likewise, an account of disagreement in general needs to preserve our intuitions about normative disagreement and normative judgement.

2. A Survey of Existing Explanations of Normative Disagreement

2.1 Some Preliminary Clarifications and Success Criteria

A distinction is made in discussions around disagreement between being in a state of disagreement, on one hand, and engaging in the activity of disagreement, on the other. The focus of this project is on the state of disagreement. This is a relationship that holds between attitudes. Being in a state of disagreement is in no way contingent on two agents voicing their disagreement with each other. Two people can hold attitudes that disagree without ever having met one another, and without being aware of the other's attitude. As Macfarlane (2014) puts it:

> 'The ancient Greeks were in disagreement with the ancient Indians about whether the bodies of the dead should be burned or buried even before Herodotus and other travelers made this disagreement known to them. Whether two people are in disagreement is a function of their first-order attitudes, not of their attitudes towards each other.' (p.119)

This eloquently explains the state sense of disagreement with which I am concerned. On top of this, it can act as a first success criteria for a theory of disagreement. If disagreement exists across countries, ages, and cultures, just by virtue of the attitudes held then it obtains in these kinds of circumstances. An ability to do so would also mean that it grounds disagreement in the attitudes that disagree, and thus fits Macfarlane's definition. Furthermore, although this project will be concerned with disagreement as a state, it is likely that the two levels of disagreement are closely linked - I am partially interested if a successful account of disagreement in the state sense helps to yield an account of the act of disagreement. If an account of disagreement in attitude helps to explain the act of disagreement, this should be seen as a strength of the account.

Next, consider unity among disagreement: 'Ordinary users of moral practice assume no distinction between moral and other types of disagreement, at least qua disagreement' (Sinclair, 2021, p.85). So, the way in which disagreement is generated needs to be uniform among different types of attitudes. This points towards the need to establish exactly which attitudes are subject to disagreement, to inform an understanding of exactly what the scope of a unified account of disagreement looks like.

In the literature, there are existing criteria that state that disagreement does not apply to certain mental states - supposition for example (Ridge, 2013, p.59). Of course, we can differ from others with respect to a range of types of attitudes, but they shouldn't necessarily put us in disagreement, or at least the same kind of disagreement as moral and belief disagreements. Disagreement, intuitively, is something distinct from mere difference. For example, I can suppose that murder is wrong, and you can suppose that it is not. These attitudes are different in a sense, but does this really put us in disagreement? Intuitively, the answer is that it does not. A successful account of disagreement needs to explain this too.

So what should come under disagreement? Naturally, beliefs are subject to disagreement, and I expect that any account that can meet the other desiderata should account for disagreement in belief without being too revisionary. Expressivists are likely to have something more interesting to say about normative disagreements. Much of my project will be spent examining what exactly grounds disagreement in these cases - why do clashes in normative attitudes constitute disagreements? So, it seems

as though our account of disagreement needs to generalise over at least moral and belief attitudes, in the name of unity. This means that it needs to distinguish these attitudes from those which don't form the basis of disagreement. The expressivist is looking for something that these attitudes share that makes disagreement possible. In the following sections, I will begin to look at what grounds disagreement. I will start by examining existing attempts at solving the problem. First, I will consider one more attitude that is subject to disagreement, and consider how a unified account might need to deal with it.

Expressivism has convinced philosophers outside of its metaethical and metanormative formulations. Credal expressivists (at least those who would explain the logical properties of credences in terms of having inconsistent credences towards the same content) believe that to think 'x is likely', is to have an attitude of high certainty towards the proposition 'x'. This is at odds with a descriptivist view, which would hold that it is a belief in the proposition 'x is likely' (Jackson and Moon, 2020, p.6). Credal expressivists would explain disagreement about the likelihood of 'x' as to holding disagreeing credal attitudes towards the proposition 'x', so they need an account of what it is for credal states to disagree. Better yet, they need to offer a unified account of disagreement that accounts for these attitudes. The need for, and difficulty in giving, an account of unified disagreement that includes credal expressivism is emphasised by Lennertz:

> 'Noncognitivists owe an explanation of how disagreement in credences is possible. And this account should also explain the sense in which disagreement in beliefs and disagreement in credences are both genuine disagreements—i.e. are explained by a general account of disagreement that has both as a type.' (2021, p.203)

I will also add, here, that it seems as though credences can disagree with beliefs. If I believe that proposition that the earth is round, I disagree with somebody who has 0.5 credence in the same proposition. This points to the need for an even more general account of disagreement. I do not take it that all normative or metaethical non-cognitivists are committed to being non-cognitivists when it comes to talk of the likelihood of propositions. Nonetheless, the ability for an account of disagreement to explain disagreement between non-cognitive credal attitudes would be seen as a positive consequence.

What criteria have been generated so far? The two strong criteria are that the account needs to be unified among belief and normative attitudes, and that it needs to explain disagreement as something that holds between, and in virtue of, the very mental states that disagree. The test for the former will be to 'zoom out' and see if accounts of disagreement preserve unity among different attitudes. The test for the latter will be to apply MacFarlane's example - do the Ancient Indians and Ancient Greeks disagree? Finally, the ability to explain disagreement in non-cognitive credal attitudes will be seen as a strength of the view, and although not a strict criterion, the ability to do so will be considered once a satisfactory account of normative disagreement has been presented.

2.2 How Have Expressivists Tried to Explain Disagreement?

In this section, I will survey some existing accounts of normative disagreement. Naturally, some of these accounts of normative disagreement make up part of a wider expressivist theory. I will not try to assess the overall success of the theories in which these accounts of disagreement appear. Rather, I will focus on their ability to explain disagreement. Interestingly, though, explaining disagreement seems to be a significant hurdle at which many of these theories seem to fall. Thus, any criticism of these theories qua disagreement is, to some extent, a criticism of the theory as a whole. Part of the motivation for

researching normative disagreement in the context of expressivism is that a proper understanding of disagreement is crucial for the success of an expressivist theory.

2.2.1 Stevenson

A natural starting point is C. L. Stevenson's account of disagreement in attitude, as it appears in his *Ethics and Language* (1944). Stevenson argued that there is a meaningful way in which we can disagree with respect to normative non-cognitive attitudes, such as those disagreements we encounter when doing normative ethics. We can 'disagree in attitude'. Attitudes are like desires under the Stevensonian view. Disagreement in attitude constitutes a disagreement about what to do. As such, they can be satisfied or not. Whether two desires can be jointly satisfied partly determines whether they disagree, under Stevenson's view. This straightforward account is bound to run into some problems, without clarifications. Stevenson rightly observes that sometimes people's attitudes can 'differ', without agreeing or disagreeing. For example, two competitors in a game of tennis may have jointly unsatisfiable desires - they each want to win, and it cannot be the case that they both win. It would be revisionary of our common-sense judgement about disagreement to claim that they disagree in virtue of these attitudes. In accordance with this, he adds the qualification: 'when at least one of them has a motive for altering or calling into question the attitude of the other' (p.3). Clearly, neither competitor would call into question the other's attitude, so Stevenson no longer gets the result that there is disagreement here. This does not get the right results either. Recall the preliminary success criterion that when two agents are in a state of disagreement, this is a result of 'their first-order attitudes, not of their attitudes towards each other' (Macfarlane, 2014, p.119). Stevenson's qualification makes the view too strong. It fails to account for agents whose attitudes disagree in spite of them having no knowledge of each other, or the other's attitude, like the Ancient Greeks and Ancient Indians. So, without the qualification, we are left with a view of disagreement as joint unsatisfiability of desires, which overgenerates disagreement. With it, the view undergenerates disagreement.

2.2.2 Blackburn

Simon Blackburn follows the Stevensonian view in appealing to the practical nature of attitudes in his book *Ruling Passions* (1998). His discussion of disagreement takes place in the context of metaethics. He rejects moral disagreement as mere verbal disagreement. That is, the meanings of ethical terms are not grounded by whatever features we think the terms right and wrong refer to (p.14). When the Utilitarian deems something 'good' they don't just mean that it maximises utility. When a Utilitarian disagrees with somebody about whether murder is bad, this is not a mere disagreement about the meaning and extension of ethical terms. Rather, it is a practical disagreement about whether to murder or not to murder. As such, disagreement and agreement between these practical attitudes is explained partly in terms of whether or not they can be satisfied. In order to avoid the worry that the original Stevensonian views faced, Blackburn argues for a different qualification to distinguish difference in attitude from disagreement in attitude. Here is an analogy that Blackburn uses:

‘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, p.69)

Blackburn likens ethical discussion to discussion about what practical attitudes to implement in ‘our house’. This is because, through ethical discussion, ‘intended direction is given to our joint practical lives and choices’ (p.69). So, Blackburn takes the thing that grounds disagreement to be a need to coordinate practical attitudes. In the case of ethics, this is a need to come up with some plan as to how we are to live together. This seems to account for disagreements between people with a need to engage in discussion about what to do. When they share some stable subject matter over which their attitudes cannot both be implemented, Blackburn’s account puts them in disagreement. But it does not seem to explain why the Ancient Greeks and Ancient Indians disagree. What is the subject matter over which they disagree?

2.2.3 Gibbard

Plan expressivism is a group of views that take our normative judgements to consist of planning states. For example, muder being forbidden could be represented as a set of plans to never murder under any circumstance. Allan Gibbard put forward the original plan expressivist approach (2003). Roughly, deciding that something is to be done, or required is to plan to do it in every relevant possible world. Deciding that something is not to be done, or forbidden, is to plan not to do it in every relevant possible world. Initially, this looks like it might struggle to explain disagreement - if I plan to murder, then how can somebody disagree with my plan? They can plan to stop me, but they might not be in any position to do so. Gibbard argues that when we disagree with somebody’s plan, we commit ourselves to an incompatible plan for what we would do in their circumstances. So if I plan to murder, you might disagree by planning to not murder if you were me. The relevant possible worlds over which we make contingency plans include ones where you find yourself in another person’s shoes, and as such, it is possible to make, and disagree with, plans about situations that others find themselves in. Normative disagreement, then, consists in incompatibilities between these planning states.

Why talk of planning for what to do in another’s circumstances? Gibbard uses this to make sense of judgements that we make about other agents, without characterising them as our desires. This can make sense of third-personal agent-relative judgements. Gibbard explains how, for example, this can account for the fact that Brutus might disagree with Caesar about whether Caesar is to go to the Senate on the Ides of March, despite them both desiring that he does so (Gibbard, 2003, pp. 68-69). Brutus sees going to the Senate as not in Caesar’s best interests. As such, he contingently plans not to go to the Senate if ever in Caesar’s circumstances. So, Caesar and Brutus have different plans for what to do in Caesar’s circumstances. This is why they disagree under Gibbard’s account, but not under Stevensonian ones. A possible objection that can be raised against Gibbard is that Caesar and Brutus’ ‘disagreement’ doesn’t seem to be anything more than a difference in plan. Their contingency plans seem incompatible on the surface, because they determine different courses of action for the person in Caesar’s shoes. But it cannot be the case that both Caesar and Brutus can be Caesar. Caesar plans to go to the senate, and Brutus plans to not go to the senate if he is in Caesar’s shoes. It cannot be the case that Brutus and Caesar both end up in exactly Caesar’s shoes, and need to decide on whether to go to the senate - Brutus’ being Caesar necessarily rules out Caesar being himself. It seems unclear where disagreement fits into the picture. In fact, it seems as though neither contingency plan rules the other out, in virtue of the fact that if Brutus

were in a position to implement his plan, then Caesar would not be, and vice versa. Gibbard, however, does hold that there is disagreement here. His argument is one from the usefulness of navigating life collectively: he argues that interpersonal contingency planning helps us to 'manage the chess of life' (p. 71). The value of sharing contingency plans, and thinking in general, puts some rational pressure on Caesar and Brutus. Caesar ought to defer some of his planning to others, because their perspectives and knowledge might aid his own planning. This is particularly applicable in this example, given that Brutus is aware of the conspiracy against Caesar, and thus has a better insight into the outcomes of his plan. So, Caesar has at least some reason to adopt Brutus' plan, seeing as shared planning is beneficial for agents. If Caesar did adopt Brutus' plan for himself (meaning that Caesar makes Brutus' contingency plan his own plan), this would involve ruling out his previous plan to go to the Senate on the Ides of March. This is how they disagree.

Gibbard's view is highly influential, although it is widely accepted that it does not quite get the right result for disagreement. Where does this original plan expressivism fail with regards to disagreement, then? As I have already explained, Gibbard's explanation of disagreement relies on a need to share plans and ideas. This need to share ideas is entirely contingent. Gibbard recognises this, which is why he defends the idea that it is often useful to do so. Michael Ridge notes that, where there is just no point in normative discussion, Gibbard would probably have to bite the bullet and claim that there is no disagreement (Ridge, 2014, p. 185). According to his view, situations where there is no rational pressure for somebody to consider two conflicting plans, just do not constitute a disagreement. Instead, this is just a kind of impasse. For example, I might have contingency plans for what to do in a great deal of situations that vary wildly from the contingency plans that an amoralist has. Presumably, most of our plans will be such that one of us could not coherently adopt the others' without changing their own plan. It would be natural, in these cases, to say that we disagree. In spite of this, I would not be under any rational pressure to reconsider my planning, or entertain the amoralist's contingency plans, and as such, Gibbard's definition would not put us in disagreement. Reducing disagreement to something that is dependent on our seeing value in sharing planning is revisionary of our common-sense judgements. It might be precisely those people whose plans I would not take into consideration, like amoralists, that I disagree with the most. Gibbard's account fails to capture this.

2.2.4 Worsnip

Alex Worsnip (2019) argues that disagreement between two attitudes occurs when those attitudes would be incoherent in the mind of an individual:

> 'Disagreement in the wide sense – final pass. You 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).' (p.12)

Worsnip's account relies on a notion of incoherence. There is no specific notion of incoherence that drives this account of disagreement, and understanding it as primitive, or by any other plausible definition, should yield the right results for disagreement (p.7). When the expressivist has a satisfactory definition of what it would take for an individual's mental state to be incoherent, they can take disagreement to be any combination of states that would lead to incoherence in a single individual. For example, the attitudes that

murder is wrong and murder is not wrong could not be held by an individual without incoherence. Worsnip's view, fundamentally, is that we can understand disagreement in mental states interpersonally, the same as we do incoherence between mental states intrapersonally, and that we can explain the former in terms of the latter.

The move from intrapersonal incoherence to interpersonal disagreement appears to be a straightforward one, but some more clarifications are needed as to the content of our normative attitudes. Worsnip argues that the interpersonal incoherence theory applies to non-centred attitudes, rather than centred ones. This means, for example, that my intention to sit down and work does not consist in an intention to sit down and work that is held by me, and thus centred on me, but a non-centred intention that I sit down and work. Suppose that I had this intention, and you had another intention, say to go for a walk, then we do not disagree. This is because, although having the intentions to simultaneously sit down and work and go for a walk would be incoherent in an individual, intending that I sit down and work, while intending that you go for a walk is not. It would be incorrect to make sense of these judgements as my desire for me to sit down and work, and my desire for you to go on a walk. Given that this resembles the Stevensonian account of attitudes, this might face the same issues that his account did. Alternatively, the existence of non-centred intentions like this could rely on the possibility of intending that somebody else does something. This can be accounted for under a Gibbard-style account of planning for what to do in somebody else's shoes, but this is likely to overgenerate disagreement (Bex-Priestley and Shemmer, 2021, p.503). Suppose that you and I are meeting to get ice cream. We agree that whoever arrives first should order on behalf of us both, that we will have the same flavour, and that we do not mind what, so the flavour can be picked arbitrarily. I form an arbitrary plan to order vanilla if I arrive first. The possible worlds in which I arrive first include those in which I am you, but you arrive first. So, I plan to order vanilla if I am you arriving first. You might have formed any other arbitrary flavour-plan, you might just intend to order chocolate. These attitudes would be incoherent in the mind of an individual. Thus, they disagree under Worsnip's account. However, they are arbitrarily chosen plans, and thus it runs against our intuition to say that you and I disagree. This gives us another test for the extensional adequacy of an account of disagreement, thus another success criterion. Does it allow for non-disagreement in these cases of arbitrarily formed intentions?

There is also one more worry for Worsnip's view. The worry is that some disagreements just might be disagreements about centred contents (Zeman, 2020, p.1664), thus these are disagreements that are not captured by Worsnip's view. This points to the need for a theory of disagreement that does not apply only to non-centred attitudes. I will not take this to be a criterion, seeing as it is probably covered by the need for unity in disagreement. As long as an account is both unified and not limited to non-centred attitudes, this should not be a worry.

2.2.5 Ridge

To reiterate the problem with Gibbard's view of disagreement, there is nothing necessary about listening to and respecting other people's plans. It only matters contingently if other people's plans for what to do in our shoes are different to our own plans. The account of 'disagreement in prescription' offered by Michael Ridge (2012, 2014) aims to overcome this worry. Ridge's approach is to continue to frame disagreement in terms of sharing normative attitudes, but does not appeal to its usefulness to ground disagreement. He departs from Gibbard's approach where, instead of positing a real, existent reason to

take other people's plans and advice seriously, he defines disagreement in terms of a counterfactual scenario in which agents both give and take advice in an idealised way. Say that I plan to see a movie. My friend, knowing that I have a lot of work to do, and my seeing a movie will ultimately leave me very stressed, thinks that I should not. Obviously, in real life, I might not feel any pressure whatsoever to adopt the plan not to see a movie for myself - my friend would not go if he were in my shoes, according to Gibbard's account of third-personal normative judgements, but that does not matter to me. However, imagining an idealised scenario where I take my friends advice seriously, I would end up with a plan to both see, and not see, a movie. This constitutes a disagreement in attitude, for Ridge, because I would be irrational if I adopted both of these plans. Of course, just as people aren't always guaranteed to take our advice seriously, we are not always guaranteed to advise others on the basis of what is in their best interest, or what they have the most reason to do, but Ridge accounts for this in his idealised scenario, too. His complete definition is:

> '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 Φ-ing in C and B would advise Ψ-ing in C, where Φ-ing and Ψ-ing are incompatible, in the sense of being impossible to combine without D's thereby being irrational.' (Ridge, 2012, p.56)

The conditions for prescription being 'honesty, full candour, and non-hypocrisy' constrain judgements that disagree as to limit the scope of disagreement. Without these conditions, I could advise a competitor in a game of poker to raise, where I know that it would be in his best interest to fold, and I would disagree with somebody who would advise him to fold, even if we both understood that folding is in my competitor's best interests. This preserves the agent-relativity of some normative judgements. This means that it can account for why, per Gibbard's example Caesar and Brutus disagree about what Caesar ought to do, despite desiring that he does the same thing. My judgement that my competitor should fold comes apart from my own desire that he raises. The former is the subject of disagreement in prescription, the latter is not. This is because the former, agent-relative judgement would be what I would advise if I were being honest and candid, and thus would apply in his counterfactual case. Not all attitudes can play an advice-giving role, even in the counterfactual scenario. This is another important part of Ridge's view.

> 'The expression of belief and desire can each in themselves constitute a form of advice. Not so for the expression of all other contentful states of mind. This, in conjunction with the proposed definition of disagreement in prescription, explains why we are perfectly happy to talk about disagreement in attitude and disagreement in belief, but not so happy to talk about disagreement in supposition, fantasy or perception.' (Ridge, 2012, p.59)

There is nothing about my expression of a supposition that serves to advise other people to adopt it, according to Ridge. When we express beliefs and desires, on the other hand, this can constitute advice for others to adopt these beliefs and desires. It is appropriate to try to guide people's beliefs and motivations by expressing these attitudes to them. So, Ridge's account draws the line between attitudes that are subject to disagreement, and those that are not. This was one of the original desiderata for an account of disagreement. Some of the other types of attitudes that are intuitively subject to disagreement should fall nicely under this account. Consider credences. We can give advice by expressing our degrees of belief. Just as we advise people on what to believe outright, we express advice in ways such as: 'it's a likely possibility', 'I wouldn't be so sure', 'you should suspend judgement'. The same can be said for specific types of desires. Normative judgements, such as moral and aesthetic ones, fall under the advising category. Expressions such as: 'no, that is not ok', or 'your shoes don't match your trousers', also prove to

be both apt for advising, and disagreement. Conversely, I don't seem to be able to coherently advise anybody on what their personal preferences are, or to fantasise about something by expressing my own preferences and fantasies.

Ridge invokes the notion of irrationality to define the type of clash in prescription that constitutes disagreement. Irrationality encompasses inconsistency in belief and inconsistency in intention. Notice, however, that not all inconsistency is disagreement. In Ridge's account, advice-apt attitudes are the only ones subject to disagreement. There exist attitudes that are subject to consistency constraints that are not subject to advising or disagreement, and this is captured by the account. The expression of suppositions do not, according to Ridge, constitute a form of advice. As such, suppositions that are inconsistent don't disagree. So, believing that grass is green and believing that grass is not green are attitudes whose expression can in themselves constitute a form of advice. They are also jointly inconsistent. As such, anybody who believed both propositions, and thus anybody who 'takes' the hypothetical advice, would have inconsistent beliefs. They would, then, be irrational in virtue of this inconsistency in belief. This is why beliefs disagree. On the other hand, according to Ridge, supposing that grass is green and supposing that grass is not green does not work in the same way. This is because the expression of suppositions do not in themselves constitute a form of advice. Even though the propositions are jointly inconsistent, the expression of these suppositions do not necessarily advise that anybody adopt these mental states.

Ridge's conditions for the type of third personal advice that are subject to disagreement risks collapsing into Gibbard's account. Imagine how the conditions would play out. Take 'non-hypocrisy'. Bex-Priestley and Shemmer (2021, pp. 7-9) worry that this might simply mean that our advice to others needs to be in line with what we would do in their circumstances. But then this account ends up with a worry similar to that faced by Gibbard. It ends up identifying our third-personal normative judgements with first-personal normative judgements about what we would do in such circumstances. If our judgements about what other people should do boil down to judgements about what we ourselves are to do in their circumstances, then why does it matter if these plans are different? Ridge could tweak the conditions so that 'non-hypocrisy' means something other than 'what we would do'. But it is unclear whether the account would be extensionally adequate. Consider Gibbard's account again. Personal contingency planning does predict the right results for agent-relative judgements, so there might not be another viable candidate for non-hypocrisy. The problem is that when we take 'non-hypocrisy' to entail Gibbard-style contingency planning, we seem to get the right characterisation of third-personal normative judgements, but fail to ground disagreement. Our judgements for others do seem to be identifiable with what we would do in their circumstances.

I think that Ridge does get something important right. The attitudes that are subject to disagreement are those that are able to play some kind of attitude-guiding role in virtue of their expression. In matters of desire and belief, we often see the formation of attitudes as a shared endeavor. Gibbard knew this, and appealed to it in his own theory. What makes Ridge's account distinct from Gibbard's is that it is the possibility of these attitudes to act as advice that grounds disagreement, and not the contingent fact that they do. It is not clear whether advice is a sufficient explanation of this role, but something like it could be. For one thing, the non-hypocrisy condition for advice in Ridge's idealised scenario seems, like Gibbard's theory, to capture the content of agent-relative judgements, albeit in a way that cannot make sense of disagreement.

2.2.6 Sinclair

I will turn to another expressivist account of disagreement. This appears in Neil Sinclair's *Practical Expressivism* (2021). The project that Sinclair undertakes is to explain the properties of moral attitudes in terms of their functions. Moral disagreement is given this treatment. So, how does Sinclair define disagreement in terms of the constitutive function of attitudes? Although the aim of *Practical Expressivism* is to explain moral attitudes, Sinclair does not only provide an account of moral disagreement. He emphasises the need for unity in a definition of disagreement (p. 85). He offers several explanations of disagreement at various levels, all encompassed by a general definition of disagreement in thought: 'Two agents disagree-in-thought iff they have mental states that cannot collectively fulfil their constitutive functions.' (p. 89). So what is the function of normative judgements? Sinclair argues that 'the constitutive function of belief is accurate representation, and the constitutive function of attitudes is to bring about the realization of their content' (p. 89). Here, it seems as though Sinclair takes the Stevensonian view of attitudes. So the function of normative judgements is to realise their content. If I have the attitude of requiring *p*, then the realisation of *p* is what the successful functioning of this attitude consists in. So, it is the impossibility of the realisation of the content of two normative judgements that matters for disagreement. The definition of disagreement seems to get the right result for explaining why forbidding and requiring the same thing are attitudes that disagree. Requiring *p* would be fulfilled if *p* were realised, and forbidding *p* would be fulfilled if *p* were not realised. It cannot be the case that *p* is true and not true, thus, requiring *p* and forbidding *p* are attitudes whose functions preclude the other's functioning - they disagree.

One upshot of this account of disagreement is that mental states of different types can disagree (Sinclair, 2021, p. 89). This is important in cases where people disagree in virtue of being in states of belief and suspension of belief, with regards to the same content. Suppose that you think that grass is green, and I suspend judgement that grass is green. This theory allows, rightly, for these attitudes to be in disagreement, as long as the constitutive function of suspension of belief is such that it cannot jointly function with beliefs. Sinclair's theory might wrongly yield disagreement in cases where, for example, somebody might believe that *p* and desire that *not p* (Köhler, 2022, pp. 517-518) This is so in virtue of the claim that the proper functioning of the desire that *not p* is the realisation of *not p* - its becoming true, and the proper functioning of the belief that *p* is *p* being true. Thus, the two mental states cannot jointly function - they disagree under this account. There is also the important worry for any Stevensonian account of disagreement, that it cannot capture agent-relative judgements without disagreement. Sinclair's account does not provide a solution to this (Toppinen, 2022, 153). This particular issue likely lies outside of the scope of 'Practical Expressivism', though. Seeing as Sinclair's aim is to explain moral attitudes in particular, perhaps he is concerned specifically with agent-neutral 'rules' around which we coordinate, rather than these normative judgements that pertain to specific individuals at certain times. The issue with agent-relativity is a general worry for Stevensonian expressivist accounts, and as such not an objection to the Practical Expressivist approach in particular. Nonetheless, it is yet to be seen how an account like Sinclair's can solve this. Sinclair's account of disagreement is, in some aspects, attractive because it explains the unity of disagreement. It explains which attitudes are subject to disagreement and why, and it does so in terms of the attitudes respective functions. For these reasons, defining disagreement in terms of the ability of attitudes to collectively function seems to be a promising approach. The worries, that of agent-relativity and wrongly yielding disagreement across different attitudes are not objections to Sinclair's definition of disagreement as the preclusion of joint functioning, but point to a need to give an

alternative expressivist account of the functions of normative attitudes. This provides more success criteria for a satisfactory account of normative disagreement.

2.2.7 Bex-Priestley and Shemmer

One expressivist-friendly approach to disagreement diverges from those I have looked at so far. One thing that the previous accounts share is that the l notion of disagreement is descriptive; if a certain set of conditions are met, disagreement exists. I will now examine a view that holds that disagreement is normative. Graham Bex-Priestley and Yonatan Shemmer present a normative account where disagreement yields when one of the agents has a reason to change their mind. Their 'Normative Theory of Disagreement' defines a set of conditions under which, to judge that at least one agent ought to change their mind is to judge that there is disagreement. The conditions are:

> '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.
> 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.' (Bex-Priestly and Shemmer, 2017, p.11)

The thing that grounds disagreement in this account is the standards of the agents' 'shared project'. The authors are neutral about where these projects come from, whether any of them are non-optional, and how they determine standards (p.13). Standard accounts of the common project that agents share in virtue of them having beliefs would be that they are aiming to accurately represent the world, or aiming at truth. This theory is not concerned with offering a theory of what type of attitudes normative judgements are either, and as such, the standards and constraints for when an agent ought to change their mind will depend on the account the expressivist takes up. The 'common project' that the authors think grounds disagreement will depend on the expressivist theory that one takes up. I take it that common projects are to Bex-Priestley and Shemmer what 'our house' is to Blackburn. It is the thing over which we need to make our attitudes converge. Expressivists like Gibbard would argue that agents are contingently engaged in the project of planning together, and the standards for the formation and retention of normative attitudes are those that would enable them to do so. As such, the normative account would correctly predict that Gibbard's account fails to yield disagreement in cases where people don't engage with this project. Sinclair would argue that, in the moral case at least, we have shared projects in virtue of our using moral language and attitudes because 'the function of moral practice is to produce, test, refine, and sustain mutually beneficial patterns of coordination through a distinctive, linguistic, means of collective problem-solving, viz. moral discussion' (p. 55). Creating and sticking to these patterns of coordination is the project that Sinclair would probably ascribe to agents using moral language. As such, they would disagree when one of them should change their mind in order to support this coordination.

The aim of my project is to explain how and why normative attitudes disagree. As such, the normative theory presented above doesn't directly present an answer to this question. In order to answer the question with which I am concerned, there needs to be an account of what the 'common project' is that we share with other agents in virtue of us thinking normatively, and what standards this project has. The idea that disagreement is at least partly normative seems to square up with our common-sense judgement. I do not think that the expressivist has any reason to reject the claim that to judge two mental states as disagreeing is to judge that, in the right set of circumstances, these judgements put normative

pressure on an agent to change their minds. This account helps to specify the focus of the project with which I am concerned. If disagreement is normative in this sense, then a theory of normative judgement should be able to locate this normativity. So, a suitable underlying question for an account of normative disagreement might be: where normative disagreement obtains, what reason is there for an agent to change their mind. It could perhaps also be phrased as a challenge: explain the theory in normative terms.

2.3 Taking Stock of the Success Criteria So Far

So, It seems like there are a few options available in explaining normative disagreement. Firstly, the expressivist can take incoherence as the concept that grounds and explains disagreement, per Worsnip. Secondly, they can posit that there is often some contingent reason why certain attitudes ought to be theoretically coherent per Gibbard. Thirdly, they can provide a counterfactual scenario where there would be a normative reason for our attitudes to be theoretically coherent, and define disagreement in terms of this per Ridge. Fourthly, they can explain why and how there is a real reason for our attitudes to be coherent per Sinclair. Finally, the expressivist can adopt the normative theory of disagreement provided by Shemmer and Bex-Priestly, according to which disagreement obtains when at least one of the agents has reason to change their mind, but apply the rules to a particular theory of normative judgement that provides the standards for the formation and retention of normative attitudes.

So, what success criteria does a successful account of normative disagreement need to meet? The preliminary success criteria was that it was able to explain disagreement in virtue of just the attitudes involved. This is to say that it can account for why the Ancient Greeks and Ancient Indians disagree, per MacFarlane's example. It also needs to come under a general theory of disagreement that applies consistently across different types of normative judgements, as well as to beliefs, and possibly other relevant attitudes. This is to say that disagreement about what to do is not a wholly unique phenomenon, it is subsumed under a general account of disagreement. Upon examining some previous attempts at explaining disagreement, some other success criteria have been generated. A successful account cannot overgenerate disagreement. Sometimes, a clash in intention is not enough to yield disagreement, and a successful account needs to accommodate for this. It needs to avoid putting arbitrarily formed intentions in disagreement. Also, Ridge's observation that the attitudes that are subject to disagreement play some kind of advising role seems to come close to explaining what is at the heart of disagreement. This is not to say that the expressivist should follow Ridge in defining disagreement in this way, but the interpersonal attitude-guiding properties of certain attitudes should be given some kind of satisfactory explanation. The expressivist should also avoid generating disagreement between a belief that *p* and a desire that *notp*, as to avoid the overgeneration worries that Sinclair's account faces. Not a criterion as such, but a guiding question, concerns why there seems to be a normative aspect to disagreement. Why is it the case that judging two people to be in disagreement seems to be a judgement that one of them ought to change their mind? It will be a good-making feature of a theory of normative disagreement, if it were able to answer this question. The most significant success criterion has only been noted briefly in this section, but will be explored in depth during the rest of this project - compatibility with agent-relativity.

2.4 A General Worry: Disagreement and Agent-Relative Judgements

Carlos Núñez (forthcoming) points out a general dilemma for expressivist theories of normative judgements He argues that whichever way they characterise normative judgements, the expressivist has to sacrifice either disagreement or agent-relativity. He argues that if normative judgements are conative states, states whose aim is to make the world match their contents, then they can explain either agent-relativity or disagreement, but not both. Which one they can explain depends on what they take the subject of agent-relative judgements to be. If the expressivist follows Gibbard in construing judgements about what other agents ought to do as judgements about what we are to do in their circumstances, then disagreement can't be explained. On the other hand, under a view where the subject of our agent-relative judgements about what other agents are to do are desires regarding them, and not ourselves, we can explain disagreement, but not agent relativity. He goes on to argue that if normative judgements are non-conative states, then they cannot explain disagreement at all. I will examine the dilemma as it pertains to conative states in this section. Núñez acknowledges the possibility of non-conative and hybrid mental states as the things that make up normative judgements, although rejects the possibility that they can account for disagreement.

Here is an example of Núñez's objection. I will start with a theory that captures agent-relativity, and show how it fails to capture disagreement. In doing so, the need for agent-relativity should be apparent. Accounting for agent-relativity was taken up by Gibbard, who argued that third-personal normative judgements are just first-personal contingency plans for what we would do if in somebody else's shoes. Gibbard argues that, for example, Brutus wanted Caesar to go to the Senate on the Ides of March, but judged that he ought to stay home. This is because Brutus plans to stay home if ever in Caesar's circumstances (2003, pp. 68-69). Gibbard's contingency plans are a plausible solution to the puzzle of what agent-relative judgements are if normative judgements are desire-like attitudes. It is an elegant account of how we weigh up the options that other people have and deliberate on what they should do. This is because it is exactly the same process as our own weighing up reasons and making plans, but done in light of imaginary scenarios where we are in other people's shoes. Recall, though, that Gibbard does not get the right results when it comes to disagreement. Imagine a disagreement between two people about whether Caesar should stay home or go to the senate. Gibbard captures their respective judgements as contingency plans that each person has regarding what they themselves would do if in Caesar's exact circumstances. This is not enough to obtain disagreement, though. Diverging contingency plans such as this can never be jointly acted-upon. Perhaps you and I differ in what we would do in Caesar's shoes. We can't both be Caesar, and as such, when we make different contingency plans, we are necessarily planning for circumstances in which the other's plan cannot be carried out anyway. Furthermore, even if Caesar had some interest in considering what others would do in his own plan-making, this is only a contingent fact. Caesar could easily pay no attention to what others would do in his shoes, and what appeared to be a disagreement would turn into an impasse. Núñez worries that this problem applies to any expressivist theory of normative judgement that can capture agent-relativity. This worry is well-founded, unless the expressivist can provide a stable subject matter for agent-relative judgements that is able to put agent-relative judgements in the kind of conflict necessary for disagreement.

On the other hand, views descended from Stevenson, like Sinclair's Practical Expressivism, that seem to be able to explain disagreement, don't account for agent-relativity. They can seemingly explain

disagreement because they ground the disagreement in one single issue. If the judgement that Caesar ought to stay home is a desire-like attitude toward him staying home, then Caesar and Brutus disagree because their judgements share a subject matter. This is unlike Gibbard's account. There are some unfortunate consequences to this view, though. The problem with taking Brutus' judgement that Caesar ought to stay home as being a desire-like attitude towards Caesar's own staying home is that it is incoherent with Brutus' actual desire that Caesar goes to the senate. This would mean that Brutus simultaneously desires that Caesar goes to the Senate, and judges that he ought to stay home. If Brutus' judgement that Caesar ought to stay home was a desire for Caesar to stay home, then Brutus would desire both Caesar's going to the senate and his staying home. Brutus, then, would disagree with himself, in virtue of something like the fact that these desires are incoherent/non-corealisable. But Brutus has a perfectly coherent set of judgements. Besides, it is intuitively clear that Brutus' thinking that Caesar should stay home does not consist in a desire that he does so. So, our agent-relative judgements regarding other people can not be desires in the Stevensonian sense, otherwise agent-relative judgements can collapse in this way.

This is the final, and most important success criteria for a theory of normative disagreement. It has to be a part of a theory or normative judgement that can make sense of agent relativity. This motivates a shift in focus to an expressivist theory of normative judgement that I have not yet examined. Alisabeth Ayars (2022) shares a similar worry with Núñez. Ayars offers a solution for the expressivist. Ayars constructs an account of normative judgement in both the first and third-personal case while being sensitive to this problem. Her view, Simple Plan Expressivism, captures the agent-relative nature of normative judgements and, as I will argue, lays the groundwork for understanding disagreement. It has been argued that Ayars' account of third-personal normative judgements fails to capture some kinds of normative judgements. This is a problem for the original view, but I will argue that it can be solved by adopting a different view of disagreement. Throughout the rest of this essay, I will adopt Ayars' version of plan-expressivism and use it to argue for an account of normative disagreement that is both well-motivated and extensionally adequate.

2.5 Ayars' Simple Plan Expressivism

2.5.1 Simple Plan Expressivism Introduced

In response to the problem of accounting for both disagreement and agent-relativity, Alisabeth Ayars presents an expressivist theory of normative judgement called 'Simple Plan Expressivism' (2022). Ayars develops Gibbard's plan expressivism, but separates 'plans' into 'decisions' and 'intentions'. Unlike planning, deciding to do something does not consist in intending to do it. Decisions are to be understood as states that function to bring about intentions. But what are decisions? Ayars argues that a decision is the state of being settled about how to constrain one's intentions:

> 'A decision is the state that concludes practical deliberation. The essential role of decision (and hence normative judgment) in our psychological economy is not to represent an independent normative fact, as the cognitivist would have it, but to settle or constrain our intentions up to arbitrary picking in a fashion responsive to our judgments about our reasons.' (2022, p.51)

Simple Plan Expressivism takes the content of normative judgements to be decisions, rather than plans. Thus, decisions are not motivating attitudes in themselves, but states that concern which motivations we allow to guide our action. Ayars vindicates this unusual-sounding concept by pointing out the possibility of delegated decision-making (pp. 52-53). This is plausibly what we do when we make agent-relative judgements. How does this move help the expressivist overcome Núñez's objection? Ayars defines third-personal normative judgements as follows:

> 'Simple Plan Expressivism: Just as my first-personal judgment that I should Φ is a decision that I Φ, my third-personal judgment that X should Φ is a decision that X Φ.' (2022, p.50)

Unlike traditional plan expressivism, under Simple Plan Expressivism, our third personal judgements are not judgements about what we would do if we were in the same situation as the subject of the judgement. Third personal judgements are decisions on behalf of the agent concerned. Disagreement in agent-relative judgements, then, is grounded in a stable subject matter. It concerns incompatible decisions for a single agent, and not two people's conflicting first-person contingency plans. This distinction between decisions and intentions is revisionary of traditional expressivist approaches, but well-motivated, partly because of its explanatory power. It is particularly well-positioned to account for disagreement, as I will argue in the following sections. Before that, I will say some more about the account.

Consider this approach in light of Núñez's objection. It seems as though Simple Plan Expressivism can explain the agent-relativity of normative judgements because they are decisions for others, made by constraining the reasons of the relevant subject of our judgement. What about disagreement in agent-relative judgements? Decisions for others are desire-like attitudes pertaining to the subject of the decision, and not to the decider (or planner, as Gibbard would have it), thus, there is a stable subject matter over which two people can disagree in agent-relative judgements.

The strength in Ayars' account is in this distinction between decisions and intentions. It explains a lot of phenomena that we can observe when it comes to planning. Consider for example, plans that fail. I might plan to exercise every day. If every day, I fail to realise this plan, because I form the akratic intention to stay inside and watch TV all day, my plan has failed. Ayars can explain this as a failure for my decision to form the relevant intention. This distinction can also explain why arbitrary intentions don't necessarily constitute plans. Arbitrary intentions do not come from decisions. That is, they are not the result of a deliberative process wherein reasons are weighed against other reasons in order to constrain possible options. So, we can form intentions against our better judgement, or while remaining indifferent to the options available to us, and these do not consist in fully-fledged plans.

I take it that reasons, under this account, are also defined by their functional role. Ayars appeals to Gibbard's account of judgements of reasons, extending it to the third-person case in the same way as she does with decisions (p. 51). To take something to be a consideration for settling on one course of action over another is to treat it as a reason. Reasons, then can be attitudes of desire or belief. I might decide to eat an ice cream, citing things such as my favourable disposition towards it, my belief that I can easily obtain one from a nearby ice cream van, and my desire to cool myself down, as reasons. These things are my reasons for deciding to get an ice cream because I have deliberated on them and weighed them in favour of my doing so. It is their role in my psychology that makes them reasons. I can deliberate in the same way when deciding what others should do. The difference with third-person reason judgements is that I am deliberating on what I take to be their agent-relative reasons, as well as agent-neutral reasons, and as such should not decide for them based on my own agent-relative reasons. I will discuss judgements

about reasons in more detail later in this essay. In particular, I will argue that we can disagree in these judgements.

2.5.2 Criticisms of Simple Plan Expressivism

In this section, I will discuss some criticisms of Ayars' view. In particular, I will explore Ayars' commitment to the corealiseability constraint, and whether this commitment precludes a satisfactory understanding of disagreement.

Recall Núñez's objection, which features again here. Núñez observed that theories of normative judgement seemed to be able to capture disagreement only at the expense of agent-relativity. Ayars' Simple Plan Expressivism can explain agent-relative judgements in such a way that they still have a stable subject matter over which they can disagree. On the face of it, this looks like it solves Núñez's dilemma. If, however, we consider the coherence constraints that apply to decisions, it seems as though Ayars' account is likely to wrongly yield incoherence and/or disagreement under certain circumstances. This is because there are a number of examples of agent-relative judgements where incoherence arises. So, it seems as though, although it may be explanatorily powerful, this account faces disagreement overgeneration worries. The Simple Plan Expressivist is not out of the woods when it comes to agent-relativity. Here is Ayars' own explanation of disagreement when we take normative judgements to be 'attitudes held directly toward others' actions', a position she calls the simple view:

> On the Simple View, my judgment that Aisha should climb is a plan that forbids Aisha not to climb; your judgment that it's not the case that she should climb is a plan that permits her not to climb. Both our plans pertain to Aisha, so disagree in the same way that conflicting first-personal plans do: my plan forbids something of Aisha that yours permits. (2024, p.87)

Ayars does not say more about how these attitudes disagree. One possibility that is not specifically endorsed by Ayars, but would follow from her view on the norms governing should-judgements, is that the non-corealiseability of these attitudes is what explains their disagreement. Consider Ayars' explanation of incoherence on her view. She argues that judgements are incoherent when they express non-corealiseable decisions. So, corealiseability could have something to do with disagreement whether it is directly relevant, or whether incoherence is relevant to disagreement. So, if we take disagreement to be either the preclusion of joint satisfaction or incoherence of decisions (recall that many of the views surveyed in section 2.2 took this form), then decisions disagree when they are jointly non-corealiseable under Ayars' view. Here is Ayars stressing the importance of corealiseability of decisions:

> 'The corealizability principle arises from the action-orienting nature of decisions. A decision's goal is to answer a practical question. These responses need to be congruent, as incompatible answers can't simultaneously guide action.' (2024, p.90)

A likely account of disagreement under Simple Plan Expressivism, then, is that where plans are non-corealiseable, they disagree. Either in virtue of their non-corealiseability directly, or in virtue of their incoherence, which in turn guarantees their non-corealiseability. Ayars could have a distinct definition of disagreement in mind. She could reject both of these definitions of disagreement, and argue for a definition that is separate from both concepts (example). I do not know if this is what she has in mind. Regardless of whether non-corealiseability is sufficient for disagreement, the implications of Ayars' corealiseability constraint on the coherence of normative judgements is controversial. I will present some

of the objections to the corealiseability constraint, and then argue that disagreement should be explained without appealing to it.

There are several objections to the corealiseability constraint. This is because many of its implications are at odds with our commonsense views about agent-relative judgements. Return to Núñez's challenge. One part of the problem was presented as follows:

> 'Agent-relativity is a pervasive feature of normative thought. The normative egoist who thinks everyone ought to do what is best for themselves; the decision theorist who thinks agents ought to maximize their own expected utility; the deontologist who thinks, for instance, that parents ought to do what is best for their own children, they all hold agent-relative normative views. Many of these views are plausible. All of them are coherent.' (forthcoming, p. 8)

For these agent-relative views to be coherent, it cannot be the case that the normative egoist, decision theorist, and the deontologist are incoherent when they make these judgements. It seems like the expressivist needs to account for these agent-relative cases. Simple Plan Expressivism already does this to an extent but arrives at an issue when those agent-relative judgements are not corealiseable. A normative egoist might be committed to the judgement that I do what's best for myself, and that they do what's best for themself. Plausibly, what is best for them and what is best for me might not always be corealiseable, but this should not render the normative egoist incoherent. Under this view, however, judgements that are non-corealiseable are incoherent. Therefore, non-corealiseable agent-relative judgements are, contrary to Núñez's claim, incoherent. Consider another case in favour of the coherence of at least some non-corealiseable decisions. I am playing poker with my friend. I judge that he ought to raise, while I judge that I ought to get him to fold. His raising and folding are non-corealiseable, but there is nothing incoherent about me making these two judgements. In fact, my getting him to fold might even be contingent on my judgement that he ought to raise. The charge against the corealiseability constraint is that it cannot account for the observation that agent-relative views seem to be able to yield non-corealiseable decisions without incoherence. This incoherence of agent-relative views is not an oversight of Ayars' view, though. She argues that it is in fact an advantage:

> 'It is an advantage of the Simple View that it can explain why judgments like J1* and J2* ring incoherent by holding that they express non-corealizable decisions' (2024, p. 92)
>
> (J1*) the judgment that Aisha should climb Kilimanjaro
> (J2*) the judgment that I should prevent Aisha's climbing Kilimanjaro.' (2024, p. 92)

It seems, then, that this is a case of conflicting intuitions on whether agent-relative, non-corealiseable judgements are incoherent at the metaethical and metanormative level. If we don't want all non-corealiseable agent-relative normative judgements to be incoherent at this level, then Núñez's challenge needs to be met. I take it that an account which avoids the controversial claim that the metaethical can bear on the normative in such a way as this account allows is desirable.

Ayars does have one solution to the agent-relativity challenge in some contexts. She argues that our judgements about what competitors in a game ought to do are actually judgements about what they should *try* to do. This way, these decisions acknowledge the non-corealiseability of, for example, two competitors each bringing about their own victory in a single game, and don't advocate for both being realised. When two people are playing tennis, for example, we judge that in the interest of playing a good

game, they should each try to win the point, and not that they each should actually win the point. The competitive nature is still preserved because each is trying to bring about their own victory. Even if we accept this view, the fact still remains that in contexts like the judgements *J1** and *J2** above, the corealiseability constraint renders them incoherent.

Consider another context in which corealiseability generates the wrong results. Margaret Shea presents yet more cases where interpersonal non-corealiseable judgements are, contrary to Ayars' account, coherent. For example, a case she calls *Lonely Aunt* (Shea, 2024):

> 'Ann's brother Bob has promised to visit their lonely great- aunt at the nursing home this weekend. But he is selfish and unreliable, and Ann knows he's booked a luxury getaway. Due to COVID restrictions, the great-aunt can only have one visitor per weekend. Bob should visit, but he won't—so Ann should.' (p. 129)

Shea argues that the judgements *Ann should visit* and *Bob should visit* are jointly coherent, despite them being non-corealiseable (it is worth noting that even if the judgements were phrased as conditionals, they would still entail non-corealiseable decisions - even in worlds where Bob does not visit, thus Ann should, it is still the case that Bob should visit - he is not off the hook). This is a problem for Ayars because, as was the case in egoism and playing poker, intuitively coherent sets of decisions are rendered incoherent because of their joint non-corealiseability. To switch back to talk of disagreement, it seems as though in this case, the two judgements disagree under Ayars' account, where they intuitively should not. This is another example of corealiseability getting the wrong result.

I have presented a number of cases where the corealiseability constraint wrongly puts attitudes in incoherence. Just in case the expressivist wishes to bite the bullet and accept Ayars' claims about the coherence of decisions, it is worthwhile to consider another controversial extension of Ayars' view. The publicity thesis, and the bases on which it could be rejected, are laid out by Ayars as follows:

> 'Publicity Thesis: If R is a reason for X to Φ, and Y's Ψ-ing would interfere with X's Φ-ing, then R is a reason for Y to not Ψ.' (2024, p.96)

> 'But if one is not sympathetic, the explanation I've presented can be interpreted as a challenge. I've illustrated that the Publicity Thesis follows from a series of reasonable theoretical choices: (1) Endorsing plan expressivism about first-personal should-judgments, over cognitivism; (2) Extending the view to third- personal judgments in a natural way, culminating in a view that improves on Gibbard's; (3) Recognizing certain rational norms governing should judgments that follow from the non-cognitivist theory; and (4) Adopting a complementary non-cognitivist view of judgments about reasons and a corresponding coherence norm governing such judgments, from which (a commitment to) the Publicity Thesis follows. Anyone skeptical of the Publicity Thesis must reject one of these steps; the challenge is to specify which.' (2024, pp.99-100)

Here, Ayars details how the corealiseability constraint, thus the publicity thesis, comes about under her view. I will argue that the expressivist should reject Ayars' approach to (3). The expressivist should not accept realiseabiltiy as the norm governing normative judgements. This is because, in doing so, they let this norm bear on which attitudes we treat as reasons, and this is how Ayars generated the *Publicity Thesis*. Why is this a problem? Part of the functional role of decisions under Ayars' definition is to be 'responsive to our judgements about reasons' (2022, p.51). I argue that decisions are not properly responsive to our judgements about reasons if we prioritise realiseability, and let this norm constrain which attitudes we can coherently judge as reasons bearing on our decision-making. The objections from Núñez and Shea, as well as Ayars' own publicity thesis, show how doing so lands the expressivist in trouble. The objection, then, is not only that the corealiseability constraint generates the wrong results for incoherence, but that it does so because it constrains decisions in such a way that they cannot be properly

responsive to our reasons. I argue that this is likely to bear on disagreement such that the corealiseability constraint and publicity thesis overgenerate disagreement. Even if they do not, the expressivist wants their account of disagreement to be compatible with agent-relativity and cases like *Lonely Aunt*. As such, the expressivist should argue for a different rational norm governing should-judgements. I will offer such an account, doing so by focussing on disagreement between normative attitudes.

2.6 Consolidating

So far, I have examined several expressivist-friendly explanations of disagreement to generate success criteria against which to test a theory of normative disagreement. I have explored a dilemma presented by Núñez and an approach that almost solves it. I have argued that, although Ayars' theory of normative judgement appears well placed to explain normative judgements, agent-relativity, and disagreement, it faces some grave objections due to the corealiseability constraint on decisions. In light of these objections, I propose that the expressivist should explain disagreement under Ayars' theory, but without appealing to the non-corealiseability of decisions. This will be the focus of the rest of this essay.

3. A 'New' Theory of Normative Disagreement

3.1 Disagreement as the Preclusion of Joint Functioning Revisited

3.1.1 Motivating and Explaining the View

So, Simple Plan Expressivism lays a lot of the groundwork for an expressivist account of normative disagreement. The problem is that it cannot make sense of non-corealiseable agent-relative judgements because of its commitment to corealiseability as the rational norm governing should-judgements. This norm wrongly construes all non-corealiseable judgements as being incoherent, where some of the cases above have shown that it should not. If this norm applies to disagreement, then it will overgenerate disagreement where it overgenerates incoherence. An account of disagreement needs to be constructed by appealing to a different rational norm. So, I disagree with Ayars' treatment of the rational norms governing should-judgements, and will opt for an alternative view in explaining disagreement. This new view avoids worries for disagreement outlined in the previous section, and shows how the expressivist has the resources to deal with agent-relativity and disagreement, when they opt for Ayars' psychological taxonomy of normative judgements, but reject the corealiseability constraint.

In defining disagreement, the expressivist should subscribe to Neil Sinclair's general account of disagreement between mental states. That is, the expressivist should define disagreement between mental states as the preclusion of their joint functioning (2021, p. 85). I will detail the account of disagreement shortly, but will first discuss what I take to be the function of decisions. I am rejecting the claim that the function of decisions is the realisation of their content, as this will generate the corealiseability constraint

under this view. Also, if decisions do not consist in motivations to bring about the realisation of their content, then their realiseability should not bear on how they stand in relations of disagreement.This does not mean that normative judgements don't play their specific role within the context of a larger system of mental states that has the more general function of bringing about the realisation of their content. Rather, I am arguing that the norms governing decisions are those that allow them to play their more subtle role in this system. What is their function, then? I appeal to Ayars' own definition of decisions:

> 'The essential role of decision (and hence normative judgment) in our psychological economy is not to represent an independent normative fact, as the cognitivist would have it, but to settle or constrain our intentions up to arbitrary picking in a fashion responsive to our judgments about our reasons.' (Ayars, 2021, p.51)

To explain the rational norms that govern should-judgements, the expressivist should appeal to the ability of the judgements to fulfill their role in guiding intentions, rather than their realiseability. They should do so because the function of these judgements consists in their successfully guiding intentions. It has already been established that, under Simple Plan Expressivism, decisions are not themselves motivational states (this is how we can decide what others should do). If decisions are understood in terms of their 'essential role […] in our psychological economy', then their ability to function depends on their ability to constrain and guide intentions.

Before detailing the view, why define disagreement in terms of functions in the first place? The case for explaining disagreement in attitude terms of the preclusion of their joint functioning can be seen by examining the merits of Sinclair's view (Section 2.2.6.). For one thing, his account of disagreement is general enough to apply to all kinds of attitudes in order to uphold the intuitive principle that disagreement is a unified phenomenon. The ability to function is broad enough to apply at least to beliefs and normative judgements, and this explains the unity of disagreement. In addition to this, Sinclair's account of disagreement provides a satisfactory explanation of why disagreement can occur across attitudes, such as that of belief and suspension of judgement. One reason that Sinclair's definition didn't provide a satisfactory solution to the aims of this project was that it lacked a way of explaining agent-relative judgements - this was due to *Practical Expressivism's* appeal to the Stevensonian view of attitudes. As such this was not a direct criticism of Sinclair's definition of disagreement, but it shows how this theory is insufficient to answer the question with which I am concerned. I will take Sinclair's definition of disagreement in mental state, and apply it to normative judgements under the Simple Plan Expressivist framework. I am not necessarily objecting to Sinclair's definition of the function of attitudes being 'the realisation of their content' outright. Perhaps, for example, the realisation of the content of decisions can be understood, not in the Stevensonian sense of realisation, but rather as the successful formation of intentions. Such an approach would be, roughly, along the lines of that suggested by Sebastian Köhler (2022, p. 518). Whether the expressivist rejects realisation as the function of normative attitudes outright, or argues that the realisation of these attitudes actually consists in their forming intentions, the theory of normative disagreement that I am concerned with can be defended under Sinclair's general definition of disagreement between mental states. In other words, whichever way you look at it, the successful functioning of normative judgements consists in their forming intentions. (This might leave the expressivist wondering why they can't just keep the corealiseability constraint but define the realisation of decisions as the successful forming of intentions. This might work under the latter reading of the function of attitudes, but the view no longer resembles that which Ayars puts forth. For simplicity, then, I will continue to refer to this move as a rejection of the corealiseability constraint.)

With this in place, I will emphasise the motivations for adopting a theory of normative judgement that accounts for agent-relative views, over the Stevensonian-style reading of the one that appears in *Practical Expressivism*. Consider Ridge's counterfactual advising account of disagreement. He emphasised the role of expression in agent-relative judgements. Recall his explanation as to why some attitudes disagree, and some do not:

> 'The expression of belief and desire can each in themselves constitute a form of advice. Not so for the expression of all other contentful states of mind. This, in conjunction with the proposed definition of disagreement in prescription, explains why we are perfectly happy to talk about disagreement in attitude and disagreement in belief, but not so happy to talk about disagreement in supposition, fantasy or perception.' (Ridge, 2012, p.59)

The ability for expressions of beliefs and desires to function as forms of advice is what makes them distinct from other attitudes in Ridge's account (although Ridge does not use talk of functions). This appears to adequately establish the uniqueness of these types of judgements that makes them subject to disagreement. The fact that beliefs and desires seem to be relevant attitudes when it comes to advising others to update their mental states, while other attitudes (Ridge uses the examples of fantasising and supposing for the sake of argument) are not, seems to be relevant for explaining which attitudes can disagree. Importantly, we can advise people based on their agent-relative reasons for acting and not on our own desire-like attitudes - recall the Caesar and Brutus case. This sets Ridge's (and, of course, Gibbard's) account apart from Stevensonian definitions of disagreement in attitude. Ultimately, Ridge's idealised conditions for advice seemed at risk of either collapsing into Gibbard's view, or not capturing our agent-relative judgements accurately. This is not to say that the approach was entirely misguided. Under Simple Plan Expressivism, the expression of agent-relative decisions for others *is* advice-giving (supposing that the decisions were made in such a way that was responsive to the deciders judgements about the agents agent-relative reasons).

Having established why explaining disagreement in terms of functions is the way forward, talking a little about what the functions of normative judgements are, and why we should opt for Ayars' view, I now will argue that the Simple Plan Expressivist theory of normative judgement, and Sinclair's definition of disagreement in mental states can explain normative disagreement. I will examine the function of normative judgements to explain how and when they disagree. Normative judgements are given the following functions under Simple Plan Expressivism:

> 'settle or constrain our intentions up to arbitrary picking in a fashion responsive to our judgments about our reasons' (Ayars, 2022, p.51)

> 'Decisions answer the practical question of what to do, and function well when they issue in an intention and thus motivate the agent to act.' (Ayars, 2022, pp. 51-52)

I take it that this means the successful functioning of a decision consists in constraining a range of possible intentions to a particular intention (or set of intentions), answering the practical question of what to do, and then motivating the agent to act in accordance with this. How can decisions preclude each other's functioning, then? I will start with what it is for decisions to malfunction. It follows from these functions that decisions malfunction when they do not constrain intentions, answer the practical question of what to do, or motivate the agent to act. So decisions preclude each other's functioning, thus disagree, when they necessarily jointly cause one or more of these malfunctions.

I argue that decisions necessarily malfunction when they function to form intentions that are inconsistent. So, what does it mean for intentions to be inconsistent? Michael Bratman (1987) argues for consistency constraints on 'plans'. Plans are motivating states made up of intentions under Bratman's view (p. 32), and thus should be understood in terms of intentions, and not decisions under the current framework:

> 'To coordinate my activities over time a plan should be, other things equal, internally consistent. Roughly, it should be possible for my entire plan to be successfully executed.' (p. 31)

According to Bratman, for an agent's plan, and therefore a collection of intentions, to be consistent it should be executable. So inconsistent intentions are those that are jointly unexecuteable. This rough definition is the one that I will proceed with. There are possibly two ways to view decisions that function to form intentions that are jointly inconsistent. On the view that I am advancing, they preclude each other's functioning, but another view might be that they can cofunction, but function to bring about an unsatisfactory result. Why take up the former view, then? One way to look at this is as follows. Under the Simple Plan Expressivist framework, the successful functioning of my decision to *x* involves constraining my intentions as to exclude intentions inconsistent with doing *x*. In virtue of this, I am precluded from constraining my intentions to include only those inconsistent with doing *x*. Each restriction of intentions precludes the other. So, decisions that function to bring about intentions that are inconsistent with one another necessarily preclude each others' functioning in this way.

In spite of this, inconsistent intentions can still motivate agents. Is this a problem for this view? Suppose that I decide to eat at a new restaurant in town, but that I also decide that I ought only to support businesses with ethical practices. Under the new view of normative disagreement, these judgements disagree. However, unaware of the goings on in the restaurant, which treats its workers poorly, I form the intention to go and dine there. The decisions seem to have successfully motivated me, and as such, don't seem to preclude each other's functioning in this respect. In spite of this, I argue that these decisions still cannot give an answer to the practical question of what to do, which is essential to their functioning. The sense in which there is no answer to this question is that there is no realiseable answer, and practical questions need answers that can be fulfilled. Recall the understanding of Bratman's rough definition of inconsistency in plan as intentions that cannot be executed. There is no realiseable answer because the intentions that these decisions function to bring about are inconsistent. There is no course of action that satisfies both of the intentions that my decisions function to bring about, and thus no practical answer to the question of what to do. The requirement for consistency in intention means that decisions are still connected to realisability, albeit not in the sense that Ayars argues for. The distinct picture of the connection between decisions and realiseability is an important aspect of this view, and will be discussed in section 3.2.6. I will briefly note the important difference here, though. Under the new view, an agent's intentions need to be corealiseable, which is distinct from the view that decisions need to be corealisable.

Before moving on, I will note one more consistency constraint on our practical reasoning, as highlighted by Bratman (1987):

> 'Roughly, it should be possible for my entire plan to be successfully executed given that my beliefs are true.' (p. 31)

Does this pertain to the functioning of decisions in the same way that intentional consistency does? Probably. It is plausible that to answer the practical question of what to do, just as an intention needs to be executable relative to the agent's other intentions, an intention also needs to be executable relative to the

state of affairs in the world that the agent is in. Intending to do something inconsistent with my beliefs seems like a failure to form a practical plan. This points to the possibility of decisions disagreeing with beliefs. You might judge that I ought to drive my car to campus. I could disagree with you just by asserting that I do not own a car. Your decision functions to bring about an intention in me that is inconsistent with my beliefs. Your decision, combined with my beliefs, necessarily fail to form a realiseable plan.

Where Gibbard's original planning theory talked of plans, Ayars' talks of reasons, decisions and intentions, defined by their distinct functional roles. The fact that decisions are not themselves motivational states makes them an apt candidate for explaining what normative judgements, particularly third-person ones, are. I think that the expressivist should adopt Ayars' taxonomy of normative judgements. That is, they should agree that they are decisions, the function of which are to constrain reasons and give rise to intentions. They should argue that successful functioning of these decisions consists in them bringing about intentions in the subject of the decision. When these judgements necessarily cannot successfully function, they disagree. To explain why decisions necessarily fail to bring about an intention, I have appealed to intentional inconsistency. This relies on the view that a pair of inconsistent intentions don't jointly constitute a practical answer to the question of what to do, seeing as they are inexecutable.

3.1.2 Exploring Some of the Features of the New View

So, my proposal is that disagreement in normative judgements can be explained in terms of the possibility of these judgements bringing about intentions. In this section, I will present some of the features of Ayars' taxonomy of normative judgements that, combined with the view of disagreement as the preclusion of joint functioning, can adequately account for disagreement in the right places. I aim to stress the importance of the gap between reasons, decisions, and intentions in obtaining disagreement in the right range of cases. In doing so, I will make sense of the departure from Ayars' treatment of the coherence constraints of decisions, particularly in the case of third-person normative judgements. The reasons for this will be made clear in the following section. It is worth starting with a straightforward example of disagreement. Take a case from Ayars:

> On the Simple View, my judgment that Aisha should climb is a plan that forbids Aisha not to climb; your judgment that it's not the case that she should climb is a plan that permits her not to climb. Both our plans pertain to Aisha, so disagree in the same way that conflicting first-personal plans do: my plan forbids something of Aisha that yours permits. (2024, p.87)

On the new view, the two decisions, one that says that Aisha should climb and one that says that she shouldn't, each function to form intentions in Aisha that, taken together, are inconsistent. Aisha can't intend both to climb and not climb, without having inconsistent intentions. This seems to get the right result and, as such, the outcome is not distinct from Ayars' non-corealiseability definition in any interesting way. It does, however, highlight the subtle difference in the approach. It is not the impossibility of Aisha both climbing and not climbing that makes these attitudes disagree, but the fact that they cannot simultaneously guide the formation of a consistent intention. So, when two or more decisions are necessarily precluded from bringing about consistent intentions, they are in disagreement. To use another example, if I judge that you ought to give more to charity, and you judge that you ought not to,

our decisions function to bring about intentions in you that are inconsistent. So, both of our decisions cannot jointly function. Importantly, it is only when a decision is necessarily precluded from forming an intention that we get disagreement, not when they simply fail to. Otherwise, I could decide that I ought to eat healthier, and if this decision fails to bring about an actual intention in me to eat healthier, I would self-disagree. This would intuitively be the wrong result.

My decision that some third person, S, does action x does not need to motivate me to bring about S's x-ing. That is to say that we are not necessarily constrained by our decisions for others. (Note that here marks an important departure from Ayars' (2024) treatment of reasons and decisions). My decision is a state of being settled about how to constrain S's options, according to her reasons. The outcome of this reasoning functions, ideally, to bring about, in S, the intention to x. The basis of my reasoning is the reasons that I judge S as having, which may well be independent and different from all of the reasons that I judge myself to have. For example, I might think that S should enter an upcoming marathon. The reasons for her doing so might be that: she has an outstanding athletic ability, she is likely to do very well, and her inevitable good performance will be good for her overall mood. None of these reasons necessarily pertain to my bringing about S's running a marathon. My decision for S does not, by itself, function to form any intention in my own mind. It is possible, for example, for me to decide to prevent S's running, without disagreeing with myself. Of course, there are situations where our reasons overlap significantly with others'. S might be a close friend, and so the reasons that she has for running the marathon, paired with me wanting the best for her, gives me reasons to decide that I should support S and try to bring about her running a marathon. More generally, the need to coordinate intentions interpersonally might mean that I should decide to act in ways conducive to S's running a marathon. This is important, and will be given a lot of attention later on. Importantly, though, our decisions for others don't necessarily constrain our own decisions and thus, our intentions.

We also don't always need to make decisions for others based on what they take their own reasons to be. We shouldn't always look to them, observe which things they typically weigh in favour of certain actions, and decide based on this. Our decisions for others are made based on our own judgements about the reasons that another has. For example, we may decide that Phil, who is an amoralist, ought not to kill. We make this decision in the absence of any reasons that Phil believes he has not to harm others. We assign this reason to him in our deciding for him. This is likely to be the case in cases of moral judgement and disagreement (I will talk more about what moral disagreements and moral reason-judgements are in section 3.3.6). Compare this to non-moral normative matters, however, such as a decision about what S should get for dinner. We are likely to be more sensitive to the reasons that S takes themself to have. So, while we sometimes assign reasons to others regardless of whether they would assign these reasons to themselves, a lot of the time we also decide on the basis of factors that apply uniquely to them. These are things like personal tastes, preferences, and tolerances. In any case, counting these as reasons consists in our own assigning them deliberative weight, and it is possible to do this without assigning the deliberative weight that other agents would assign.

The properties of Simple Plan Expressivism that I have explored make it apt as a theory of normative judgement that can account for disagreement, when combined with Sinclair's definition of disagreement as the preclusion of joint functioning. In the following section, I will examine this view in light of the success criteria that I generated in sections 2.2-2.4.

3.2 Does This Account Meet the Success Criteria?

Recall the success criteria. A successful account of disagreement should explain disagreement between attitudes in terms of the attitudes themselves, and not an agent's second-order attitudes - why do the Ancient Greeks and Ancient Indians disagree? The account should offer an account of third-person agent-relative judgements that successfully generates disagreement. The account should also avoid overgenerating disagreement, per the objections to Worsnip, Sinclair, and, most importantly, Ayars. I also noted that it is of interest whether the account can make sense of disagreement as a normative phenomenon. In this section, I will explore how this account meets these respective criteria.

3.2.1 Unity

As noted in section 3.1.1, defining normative disagreement under the general account of disagreement in thought as the preclusion of joint functioning means that the expressivist can preserve unity. So, disagreement in belief also comes under Sinclair's general account of disagreement in thought. Beliefs necessarily cannot jointly function when they cannot both be representationally accurate. This occurs when beliefs are jointly inconsistent. Given this definition, and the account of normative disagreement that I am defending, it seems as though this definition of disagreement in thought applies in a unified way to both normative and belief attitudes. I also argue for another type of normative disagreement that comes under this unified definition in section 3.3.4.

What about disagreement among attitudes? In section 3.3.1, I mention one of Bratman's (1987) rationality requirements on intentions being consistency with the agent's beliefs. The sense in which intentions and beliefs need to be consistent is that a plan should be executable relative to an agent's beliefs (p.31). Where intentions are not consistent with an agent's beliefs, the agent cannot be said to have a practical, executable plan. So, a decision can disagree with a belief where the intention that the decision functions to bring about is inconsistent with the belief in this sense. As I note in 3.3.1, a plausible sense in which two agents could disagree in thought could be as follows: you decide that I should drive my car to campus, I believe that I do not own a car. Your decision functions to bring about an intention in me that, combined with my belief, cannot form an executable plan. This is distinct from having an attitude of believing something while thinking that it should not be the case (see section 3.2.4).

Sinclair acknowledges that his account has the upshot of being able to explain why attitudes of belief and belief-suspension towards the same content disagree (2021, p.89). In order to do so, however, there needs to be an account of the functions of suspending belief. Along with this comes the need to explain what the functions of attitudes of degrees of belief are. In section 2.1, I mention that expressivists of the kind I am concerned with in this essay could also be sympathetic to the idea that credences are attitudes of certainty that we hold towards content, and not beliefs about certainty. Presumably, they would say the same thing about suspension of belief. So, can this account of disagreement account for belief-suspension disagreements and disagreement in degrees of belief? If the functions of attitudes of (un)certainty are such that they can preclude each other, and beliefs, from functioning, then this account will succeed. Ultimately, the credal expressivist would probably be able to provide such an account, in virtue of their explanation of what attitudes of non-belief and degrees of belief are. This account would need to allow for degrees of belief to preclude the functioning of beliefs. If they can do so, then this general account of disagreement can be unified with belief-suspension and degree-of-belief disagreement.

3.2.2 The Ancient Greeks and Ancient Indians

> ‘The ancient Greeks were in disagreement with the ancient Indians about whether the bodies of the dead should be burned or buried even before Herodotus and other travelers made this disagreement known to them. Whether two people are in disagreement is a function of their first-order attitudes, not of their attitudes towards each other.’ (p.119)

They disagree about what to do with the bodies of the dead in virtue of their decisions that preclude each other’s functioning. The Ancient Greeks’ judgement that the bodies of the dead ought to be buried constitutes a decision that everybody buries their dead, while the Ancient Indians’ judgement constitutes a decision that everybody burns the bodies of the dead. Nobody can intend to both bury and burn the bodies of the dead (assuming, of course, that these are non-corealiseable) without inconsistency. So, these decisions cannot jointly function. This relies on the possibility that in the same way as we can decide for others, we can decide for everybody. That is, we can make universal normative judgements that function to bring about an intention in everybody. If this is the case whenever we make normative judgements that are agent-neutral, then this view can explain why the Ancient Greeks and Ancient Indians disagree.

3.2.3 Agent-Relativity and Undergeneration

Ayars’ theory of normative judgement gets the expressivist off the hook for the worries facing Gibbard and Ridge. We can explain what these views succeeded, and failed, in capturing when it comes to normative disagreements. Take the failure of Gibbard’s view to explain disagreement in cases where there is no point in exchanging plans, and the worry that Ridge might inherit these problems. If our third-personal judgements are first-personal plans for what to do in another’s shoes, the account fails to generate disagreement in cases where there is no need to consider other people’s plans. The account is missing something. This is not a worry for the expressivist who defines our normative judgements as decisions - another agent can be the subject of a decision, so our third-personal decisions don't concern ourselves. Decisions are desire-like in the right way to explain disagreement - they concern another’s actions. Not only this, but the expressivist can account for disagreement while explaining why our judgements for others resemble first-person contingency plans or non-hypocritical advice. There seems to be a symmetry between these judgements that Gibbard and Ridge successfully capture. This can explained under Ayars’ taxonomy by saying that the way that agents deliberate on reasons and make decisions is going to be the same whether those reasons belong to themselves, or to another. Under the new view, the expressivist does not identify first-person contingency plans with third-personal decisions, because the subjects that these decisions function to bring about an intention in are distinct. However, the decisions we make for others, based on their reasons, are likely to be the same as the decisions we would make for ourselves, if their reasons were our own.

So, the expressivist does not need to explain why other agents' contingency plans are worth considering, instead, they can appeal to the function of decisions being to bring about intentions, not just in ourselves, but in other agents. Thus, they can explain why these attitudes disagree without appealing to their usefulness of shared planning. This means that they avoid the worry of undergenerating disagreement in circumstances where agents don't jointly plan. The expressivist can also explain the connection between advice and judgements. Recall that the worry facing Ridge was that if our advice

reduces to what we would do in another's shoes, his view collapses into Gibbard's, thus he faces the undergeneration worry, and if it does not, it is unclear in what sense the advice is 'non-hypocritical'. Offering genuine advice might just be the expression of an agent-relative decision for another, made upon considering their subjective (and objective) reasons. Thus, this view captures how the attitudes that are subject to disagreement seem to be those that we would advise for others, and that we would adopt if in their shoes. It does so because it takes up the Simple Plan Expressivist framework.

3.2.4 Overgeneration Worries

One of the success criteria was generated from concerns pertaining to Sinclair's account. This worry was that it puts the belief that *p* and the attitude of wanting that *not p* in disagreement because the realisation of *not p* precludes the truth of *p* (Köhler, 2022, pp. 517-518). Sinclair's account got the wrong results because of the combination of the definition of disagreement in terms of preclusion of joint functioning, and the definition of the function of attitudes being the realisation of their content. How does the account that I have been endorsing handle this case? If the function of the desire that *not p* is to bring about an intention, then this attitude's successful functioning does not preclude the truth of *p*. If a normative judgement that *p* should be the case functions to bring about an intention to make *p* true, and as such the successful functioning of that judgement doesn't consist in *p* actually becoming true, then *p* does not need to be realised for the attitude to function. This means that both mental states can successfully function. On this account, they do not disagree. Imagine that I am entering a race. I decide that I should win, even though I do not believe that I will. I can still form an executable plan to win, relative to my beliefs, even if the plan will likely fail to be realised (this relies on the premise that there is a sense in which my winning is possible - if, on the other hand, I believe that I will not win because I have not entered the race and will not take part, then there probably is disagreement of the type explained in section 3.2.1). So, rejecting the claim that the function of desire-like attitudes is to bring about the realisation of their content rescues the definition of disagreement from overgenerating disagreement in these cases. The rejection of the idea that normative judgements function when their contents are realised also has the advantage of explaining cases involving agent-relativity, which I will explain in detail later in this essay.

A successful account of disagreement also needs to avoid the objection to Worsnip's account. The objection was that it wrongly put some cases of arbitrarily formed intentions in disagreement, thus overgenerating disagreement. Under Ayars' distinction, when intentions are formed arbitrarily, they are not the result of decisions. Given this distinction, and the theory of disagreement as the preclusion of joint functioning, if you and I have intentions that cannot jointly be realised, this does not constitute disagreement. Our intentions are private, and there is no necessary rational pressure to make our intentions corealiseable with the intentions of others. This is why two competitors in a game of tennis do not disagree when their intentions are non-corealiseable. To illustrate, consider another way in which intentional consistency comes apart from disagreement. The aim of this example is to show that our actual intentions do not determine whether we disagree, rather, decisions do so in virtue of the intentions that they function to bring about. Suppose I go out with a friend for the night and he decides that we should go to a party, and I decide that we should not, we disagree. This is a real disagreement regardless of whether, deep down, I actually intend to go to the party against my better judgement. I don't secretly decide that I should, rather, my decision that I should refrain fails to function to bring about an intention not to go. Our

decisions function to form the appropriate intentions - that of either going or not going. Even though our intentions could be in alignment, we both intend to go, we make decisions whose functions preclude each other. One eliminates the intentions consistent with going to a party, one eliminates the intentions inconsistent with going to a party. It doesn't matter that my decision actually fails to fulfill its function. We get this result because these decisions are precluded from jointly functioning. We disagree in normative judgement, despite the consistency in our intentions.

3.2.5 Does it Work as a Normative Theory?

Recall Bex-Priestley and Shemmer's *Normative Theory of Disagreement*:

> '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.
> 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.' (Bex-Priestly and Shemmer, 2017, p.11)

If judging that two agents disagree consists in judging that one has reason to change their mind, it is of interest whether a theory of normative judgement can explain this. If the expressivist understands normative judgements as decisions, and the functioning of decisions to be the constraining and formation of intentions, then the 'project' of normative judgements is to be defined in terms of the successful constraining and formation of intentions in an agent. So the 'standards for the formation and retention of attitudes' (2017, p. 11) are those explored in section 3.1.1, that are necessary for the functioning of decisions. These standards bear on whether normative judgements disagree. This is how the view that I am defending can explain the first two conditions. But they don't give us a satisfactory idea of why judging that two people disagree is to judge that one of them has a reason to change their mind.

For this view to work as a theory of disagreement at the normative level, the expressivist would need to take on the first-order normative view that when two decisions cannot jointly function, this gives somebody a reason to change their mind. This means that they need to take on the first-order normative commitment to weighing the belief that two decisions preclude each other's functioning in favour of deciding that somebody should change their mind. To apply the normative theory to Sinclair's definition of disagreement more generally, they should argue that when two people have attitudes that preclude each other's functioning, at least one of them has a reason to change their mind. So, this theory of disagreement can work on the normative level, if the expressivist were happy to accept this first-order norm.

3.2.6 Agent-Relativity, Non-corealiseability, and Overgeneration

The most important success criteria that I established in my discussion of the literature on disagreement was capturing agent-relativity. Specifically, it should be able to make sense of non-corealiseable agent-relative judgements, and thus avoid the worries that the corealiseability constraint generated. Recall Núñez's challenge for expressivism. He argues that the expressivist should be able to make sense of agent-relativity and disagreement. Ayars' theory of normative judgement gives the expressivist an account of what third-personal agent-relative judgements are such that they can disagree, but renders

combinations of decisions incoherent when they are jointly non-corealiseable. Any account of disagreement that depends on the non-corealiseability of normative judgements will overgenerate disagreement in these cases, too. To avoid overgeneration, and committing themselves to a 'publicity thesis', I argued that the expressivist needs an account of disagreement that does not appeal to the realiseability of decisions. I have just detailed such an account, so I will now test it against the cases for which corealiseability generated the wrong results.

How does the new view of normative disagreement avoid generating disagreement and incoherence in these cases? Take the egoist. When the normative egoist says that every agent ought to maximise their own utility, he knows that the world cannot be such that everyone successfully maximises their own utility unencumbered by others' maximising of their utilities. This is to say that the decisions that he makes for everyone are non-corealiseable. But, this decision functions when it forms particular intentions in the relevant agents and settles the question of what to do. The proper and successful functioning of the normative egoist's decision would be the forming of an intention in me to do what is best for myself, an intention in you to do what is best for yourself, and so on. This function can be realised. These are consistent intentions for independent agents to have. My intending to do what is best for myself is a consistent intention to have. If what is best for me is at odds with what is best for you, it does not mean that the normative egoist is incoherent, or that he self-disagrees in any sense. His decision can still successfully function - each of our intentions are coherent and consistent in our own minds. So, the normative egoist's judgement that everybody ought to do what is best for themselves is a collection of non-corealiseable decisions, however, all of the decisions are still able to function, that is, they can form the relevant, consistent intentions in the relevant agents. When disagreement is understood as the preclusion of the proper functioning of decisions, and not as the non-corealiseability of decisions, we can preserve agent-relativity by appealing to the consistency of the intentions that the decisions function to bring about.

Take another example, Shea's *Lonely Aunt* case, where two Bob should visit his Aunt, but he won't, so Ann should visit instead. It is the case both that Bob should visit and that Ann should visit. The decisions that Ann visits and that Bob visits are non-corealiseable, but seem to be coherent and not in disagreement. Even if Bob is guaranteed not to visit, and as such it is the case that Ann should visit, it is still the case that Bob should visit (As Shea puts it, he is not off the hook). Under the view of normative disagreement that I have been defending, the realiseability of decisions does not directly determine whether decisions are in coherence or agreement. Both of these decisions can function to bring about the relevant intentions in both Ann and Bob. Deciding that Ann should visit functions to bring about an intention in Ann that she visits, and deciding that Bob visits functions to bring about an intention in Bob that he visits. Nothing precludes them from both having these intentions under this view, even if these intentions would be non-corealiseable. This is the right result. It explains how, even if Ann visits, Bob is still blameworthy, it is still the case that he should visit. I have shown how this new account of normative disagreement preserves the agreement and coherence of these non-corealiseable normative judgements.

3.3 Objections and Replies

3.3.1 Objection: Corealiseability Matters

The opponent of this view might argue that most of the time, corealiseability does bear on our normative judgements. That is, they might agree with Ayars' intuition that non-corealiseable decisions are incoherent, at least some of the time. It could be argued that the expressivist cannot just reject an important part of our moral and social lives to vindicate their theory in light of the objections presented in this essay. Under normal circumstances, there just is something wrong with making non-corealiseable decisions, and the account given does not have anything to say about this yet. In day to day life, we do seem to respect something like the corealiseability constraint in order to get along with one another and coordinate our actions. On top of this, if we observe the consistency constraint on intentions, then any decisions made on behalf of an individual need to be such that they can all be acted on. Intentions in the mind of one person need to be corealiseable, surely this sometimes applies interpersonally, too. Is it really the case that this only holds intrapersonally and not interpersonally?

The expressivist could reply that there are good reasons to reject many agent-relative normative views, a lot of these will be reasons found at the normative level, and not the metaethical or metanormative one. For example, egoism might not take into account important reasons weighing in favour of compromising and cooperating with other agents. These will be reasons to be weighed in favour when making decisions for ourselves and others, and not rational requirements for making decisions. Most of our reasons to be weighed in favour of actions when we make decisions are likely to be sensitive to other agents and their reasons. Ignoring these would be a normative failure. The reason that some judgements 'ring incoherent', to Ayars could be down to the fact that it seems like these judgements are a result of somebody poorly weighing up reasons, and ignoring the agent-neutral reasons that they have. It is an example of bad decision making, and not metanormative incoherence. We disagree with views like this because we make different reason-judgements that would result in ourselves making different decisions.

Another possibility, however, is that there are occasions where judgements like these do disagree, but that they do so because of something other than their joint non-corealiseability. This would explain why Ayars argued that some non-corealiseable decisions ring incoherent. According to one of her examples, I cannot coherently decide that somebody else should order the last burrito available at a restaurant, while deciding, myself, to order it. I have already appealed to aspects of Bratman's theory of intentions to explain why consistency in intention is part of the function of decisions in virtue of the rational requirements of intentions. I will appeal to Bratman's theory of shared intentions to explain the need for some of our decisions to be not just consistent intrapersonally, but interpersonally. The aim of the next section will be to integrate Bratman's theory of shared intentions into Ayars' theory of normative judgement and this theory of disagreement.

3.3.2 Reply: Shared Intentions

Given the existence of coherent non-corealiseable agent-relative judgements, how do we explain the observation that a lot of the time, our intentions do need to be jointly realiseable with the intentions of others? The new puzzle is this - is it possible that the attitudes in Ayars' burrito example constitutes disagreement in spite of the non-disagreement of the egoist's attitudes? More importantly, how could this be so? My answer is that Ayars would be right to spot incoherence in the burrito example, but only under particular conditions - coordination. The consistency of intentions interpersonally matters if we are concerned with aligning our intentions with others. It is plausible that when dining together, for example, when one person at the table makes a decision, it concerns everybody that they are dining with too. To explain this, I will appeal to Bratman's theory of shared intentions. In case the need for the move to shared intentions does not seem well-motivated enough, it is worth considering an issue that Maraget Shea outlines for plan-expressivists (Shea, 2024). Briefly, 'clumped judgements' are normative judgements that apply to more than one agent that are distinct from a conjunction of judgements about what each agent is to do. In a footnote, Shea points to one way in which Simple Plan Expressivism can deal with this distinction:

> 'Although they face other grave objections, two rivals of Gibbard's plan-expressivism look comparatively well-positioned to deal with the semantic distinction between clumped judgments and the conjunction of atomic judgments. The first is Ayars' version of plan-expressivism. Arguably, making decisions on behalf of a set of agents, A, B, and C, is different from making decisions on behalf of A, and separately, on behalf of B, and separately, on behalf of C.' (p. 147)

Just as normative judgements can come apart, recall Shea's *Lonely Aunt* case, discussed in section 2.5.2, judgements can 'clump' together. For example, consider the *Slice and Patch* case (Mellor and Shea (2024, p. 706) appeal to this case as an example of 'joint oughts', it originally appears in Estlund (2020), I have simplified it somewhat). Imagine a patient in a hospital for whom the best case scenario is that their doctors, Slice and Patch, do their procedures: slicing and patching, respectively. If Slice sliced without patch patching, or vice versa, the patient would die a slow and agonising death. If neither did their respective jobs, then the patient would not be healed, but they would be better off than they would be having received one half of the operation. It seems to be the case that 'Slice and Patch ought to slice and patch' is true, while it is not the case that 'Slice ought to slice' and 'Patch ought to patch' are true independently. The truth value of the 'clumped' judgement is different to the truth values of the 'unclumped' ones. Just as Shea observed with cases like *Lonely Aunt*, this case shows that there is a semantic difference between 'clumped' and 'unclumped' judgements. To explain this, I will appeal to Michael Bratman's theory of shared intentions (2014). I argue that the difference between the two types of judgements is that the former, but not the latter, functions to bring about shared intentions. Decisions made on behalf of a group of agents function to bring about shared intentions among those agents. If intentions are shared then they need to be consistent across agents. Bratman lays out the the conditions for shared intentions as follows:

> 'A. Intention condition: We each have intentions that we J; and we each intend that we J by way of each of our intentions that we] J (so there is interlocking and reflexivity) and by way of relevant mutual responsiveness in sub-plan and action, and so by way of sub-plans that mesh.

B. Belief condition: We each believe that if the intentions of each in favor of our J·ing persist, we will J by way of those intentions and relevant mutual responsiveness in sub-plan and action; and we each believe that there is interdependence in persistence of those intentions of each in favor of our J-ing.
C. Interdependence condition: There is interdependence in persistence of the intentions of each in favor of our J-ing.
D. Common knowledge condition: It is common knowledge that A-D.' (2014, p.103)

So, how should we make sense of shared intentions under the simple plan expressivist framework? Firstly, these are still individual intentions. They don't differ in this respect from the ordinary intentions we have been discussing so far. This is just to say that there is no 'we' agent that holds these shared intentions, the intentions belong to the individuals respectively. Secondly, they can come about as a result of decisions. Just as Ayars argued that decisions apply to others as well as ourselves, it can be the case that, in the same way that we decide for individuals, we decide for groups. Under the simple plan expressivist taxonomy, then, to utter 'we should see a movie' is a decision that functions to form a shared intention in two or more agents to see a movie. The intention is one that is sensitive to the intentions of other parties involved in the ways that Bratman outlines, this is because there is 'interlocking and reflexivity'. The decision 'we should see a movie' concerns the intentions of the group 'we', and not merely the intentions of the individuals that make up the group. This is just to say that the decision 'we should see a movie', is distinct from a conjunction of the decisions 'I should see a movie', 'you should see a movie'. If I told you 'we should see a movie', but then formed the intention to not see a movie, my intentions would be inconsistent. My intention not to see a movie would not only be inconsistent with my intention that we see a movie, but with your intention that we see a movie. It would make sense for you to abandon your intention that we see a movie, given the properties of shared intentions This is because your intention was responsive to my intention that we see a movie.

I will use something similar to the common car sharing case to illustrate how disagreement concerning matters of shared intentions might come about. If I want to get a taxi somewhere and, upon talking to you, discover that you also need to get a taxi to the same place at the same time, we might decide that we ought to share a taxi. Presumably, we can say that I have reasons that overlap with your reasons. I need to get a taxi, so do you. We might have a particular reason to share a taxi, seeing as we would each save money. On the basis of our overlap in reasons, one of us could utter 'we should share a taxi'. This constitutes decisions, both a first and third-personal one, that each of us should form a shared intention to get a taxi. Say that it is successful, and we form this shared intention. Now we are concerned with the details of getting a taxi. Now, our decisions need to function to form intentions consistent with our sharing a taxi. If I will only get into an Uber, and you refuse to do this because you will only get a black cab, we disagree because we each decide on something that would inhibit our sharing a taxi. My decision to get an Uber entails that we will both get an Uber, because we are sharing a taxi. Yours entails that we both get a black cab. Hence, we disagree. Our picking a particular taxi is our forming an intention under the intention that we share a taxi, this is what Bratman calls 'sub-plans'. The need for them to 'mesh' is captured by the need for consistency in intention. This appeal to shared intentions explains the difference between important cases of disagreement and difference. If I am standing waiting for a taxi to a certain place, and you are waiting for one to take you in the opposite direction, we do not need to decide on any particular means of getting a taxi. I can open my phone and order an Uber, and you can hail a black cab. We do not disagree. If, as illustrated above, we have a joint intention to get a taxi together, then our making different decisions about the particular taxi matters - this is to say that it can ground disagreement in these cases. (Of course, if we were discussing the best means of getting to our destinations, I could say that you ought to use Uber to do so, this would be the simple case of a third

person decision that functions to form an intention in you to get an Uber. I have already shown how there can be disagreement in these types of cases.)

Recall one of the motivations for this account. Margaret Shea had suggested that Ayars' account might be able to deal with the phenomenon of normative clumping. In her critique of plan-expressivism, she argues that there is 'a distinction between (i) the conjunction of the judgements that A should φ and that B should φ and (ii) the judgement that φ-ing should be done jointly—that is, by the pair of A and B' (2024, 147). In this section, I have appealed to Bratman's account of shared intentions to distinguish between these judgements. In the case of what should be done jointly, we make decisions that function to form shared intentions across agents. Shared intentions are interdependent, and as such, they need to be consistent with each other. In cases of decisions for individuals, consistency in intention only matters intrapersonally, whereas when we are concerned with decisions that function to form intentions that are interdependent across people, interpersonal inconsistency acts as a constraint. So, Shea's observation that some judgements 'clump' can be explained by the possibility that decisions can function to form shared intentions. To return to another example used, the decision that 'Slice and Patch ought to slice and patch' functions to form the shared intention in Slice and Patch to slice and patch. Suppose that this decision successfully functions, then Slice's intention to slice is interdependent with Patch's intention to patch. Their intentions pertain to the group action 'Slice slicing and Patch patching', and not just to their individual respective actions.

In order to say more about the role of shared intentions in our everyday lives, return to Ayars' example, if I am dining with somebody, it is plausible that I make decisions about what *we* are to eat. Seeing as a part of eating together is coordinating and facilitating our respective attitudes about what to eat, our decisions do not just guide our own intentions, but the intentions of the group, in virtue of the fact that our goal is to dine together. My decision to have the last burrito is a decision on behalf of everyone else to let me have it. Imagine that I expressed my decision to have the last Burrito vocally. It seems as though this would function to guide my dining partners' intentions towards themselves not getting it. It is the nature of this decision that it functions to guide the intentions of other people, too, based on a mutual understanding that we are dining together. Of course, Bratman's conditions don't need to actually be met for this to be the case. This is so by virtue of the fact that disagreement concerns whether decisions *can* form consistent intentions, and not whether they *do*. It just needs to be the case that the ideal realisation of this decision would be the formation of this shared intention. Cases such as this are likely to appear all of the time in our day-to-day lives. We are constantly involved in facilitating each other in our plans and decisions, and coordinating with each other, and this involves considering the intentions of groups of people, and not just individual agents. Given the existence and prevalence of shared intentions, and the definition of disagreement given so far, some kinds of disagreement arise when we make decisions that preclude the formation of consistent shared intentions.

So, how does this answer the objection presented above? Recall that the opponent of this view might be sceptical about throwing away corealiseability, seeing as it does seem to be a key part of decision-making in our shared lives. Instead of conceding that corealiseability is, in itself, a rational constraint on decisions, shared intentions explains the intuition behind this assumption. Shared intentions offer an account of how the compatibility of our own decisions with those of others is important without appealing to corealiseability. Assuming that Bratman's theory of shared intentions can successfully generalise over groups bigger than just two agents (2014, p. 85), the expressivist can enrich their account of normative disagreement. This is important because we often make judgements that concern the future of entire groups of people, even if they appear to concern just our own intentions. This could be the case

in Ayars' burrito example. In making these decisions, we try to coordinate the interdependent intentions of multiple people in the way that Bratman outlines. The natural way to characterise these judgements, then, is as those decisions that function to bring about shared intentions. When decisions function to bring about shared intentions, intentional consistency becomes a constraint interpersonally. This is to say that it is not just the internal consistency of the intentions that decisions function to bring about that determines whether decisions are in disagreement, but consistency across multiple people's intentions. Importantly, though, this interpersonal consistency requirement arises because we are trying to form shared intentions, and not just as a result of the coherence constraints of decisions.

One objection to this account may be that there is no clear way of distinguishing these decisions, and their distinct functional roles, from decisions that lack a coordination aspect. We can phrase this concern as a question: do we always know when our decisions function to form shared intentions? It may well be that this is implicit in our decision-making. For one way in which this question can be pushed using an example: why is there this difference between the egoist and burrito case? I have argued that deciding that somebody does some action doesn't necessarily constitute a decision for everybody else to allow them to. This is why the egoist's view is coherent. But isn't the burrito case just another example of this? I argue that there is likely a relevant psychological difference between these decisions that accounts for why one attitude functions to bring about an individual intention and one functions to bring about a shared intention for coordination. The aims of this section have been to spell out how these decisions work, and why we can sometimes be constrained by other's intentions. I have appealed to the intuitive possibility of making decisions that concern the formation of shared intentions in cases such as riding or dining together.

3.3.3 Objection: Not all Normative Disagreements can be Characterised as Disagreements in Decision

Another possible objection to the view defended in this essay is that normative judgements aren't all simply judgements about what to do. It at least seems possible to make a normative judgement about something, without being committed to a fully-fledged decision regarding it. I could say that it would be entertaining to trip my friend up as he walks past me in the street, even if I am committed to the decision to not trip him up. You could plausibly disagree with me, saying something like 'no, it would not be entertaining'. How would we disagree? We both decide that I shouldn't trip him up, it's just the case that I think that there would be some entertainment value in me doing so, and you don't. We do not disagree in decision, but we disagree over the normative value attached to my tripping somebody up. So how should the expressivist make sense of this disagreement? I will present an account of another type of normative disagreement that vindicates the account given in this essay against the intuition that there is disagreement in this case. This account does not involve any extra commitments for the expressivist, beyond what I have already subscribed to.

3.3.4 Reply: Disagreement in Reason-judgements

It is a consequence of Simple Plan Expressivism's functionalist definition of judgements about reasons, combined with the definition of disagreement as the preclusion of attitudes' joint functioning, that we can disagree in our judgements about reasons. I have taken the definition of judgements about reasons to be a functionalist one because it appeals to the role that judgements about reasons play in decision-making. Their role consists in giving deliberative weight in favour of one decision, over another. Consider an example where disagreement in reason judgements is likely to be rife - preliminary disagreements. It is often the case that we need to establish which kinds of reasons we are weighing in favour of actions in order to resolve preliminary disagreements. Imagine a disagreement between Bob and Bill as to where they should take a trip. Bob wants to go to Paris, Bill wants to go to London. The disagreement can be characterised, simply, as the preclusion of their decisions both functioning. They can't both form the shared intention to go to Paris, and the shared intention to go to London simultaneously without inconsistency. This is the standard account of disagreement already outlined. Consider however, that seeing as they want to resolve this disagreement, they might start to discuss their judgements about reasons pertaining to their decisions. Bob might cite the reason that they both have family living in Paris, Bill might cite the reason that they both have a lot of friends living in London. They discover that Bob does not give any weight to the friends consideration, and Bill does not give any weight to the family consideration. Perhaps they even count these as considerations in favour of not visiting one of the cities. Upon realising this, it seems like it would not be fruitful to carry on arguing about where to visit. Instead, they should settle the preliminary disagreement over how much deliberative weight to assign to seeing family, and to seeing friends. This might take the form of something like:

Bob: 'seeing family is an important consideration, seeing friends is not'
Bill: 'no, seeing family is not an important consideration, but seeing friends is'

Each of them urges the other to assign more deliberative weight to the other option. Bob thinks that in deciding what they should do, they should give more deliberative weight to family than friends, which is incompatible with giving more deliberative weight to friends than family. If the function of reason-judgements is to give deliberative weight to certain options to guide the formation of a decision, then these judgements disagree. The two judgements about reasons cannot jointly function to contribute to the formation of a decision, seeing as each precludes the other from weighing in favour of a particular decision. Hence, they disagree in their assigning weight to reasons. This is a disagreement in reason-judgement. Why take this to be a preliminary disagreement in itself, and not just a part of the original disagreement in decision? Firstly, this is still a disagreement pertaining to where to take the trip. It can be seen as a sub-disagreement about how the trip disagreement should be settled. Bob and Bill are engaged in trying to establish enough common ground so that they can effectively make a decision regarding the location of their trip. Imagine neither Bob or Bill was concerned with taking a trip at all. This disagreement would no longer seem to matter. Assuming, though, that each still had their respective attitudes about their reasons, the disagreement would still exist in the sense that they have attitudes that are precluded from jointly functioning.

Why not characterise this particular preliminary disagreement as disagreement in decision? Recall that decisions disagree in virtue of the inconsistency of the intentions that they function to bring about. It initially seems possible that Bob and Bill disagree because each has made decisions that function to bring

about different underlying intentions about the purpose of their trip. Bob has decided that they should see family, Bill that they should see friends. But how would these underlying intentions figure in the decision about where to take the trip? Firstly, intentions don't play a role in decision-making in quite the same way that our judgements about reasons do. Bratman (1987) argues that ordinary desires and beliefs that can act as reasons in our decision-making are distinct from intentions, which can not do so. Our prior intentions provide constraints on which decisions can be rationally accepted. It does not make sense to say that we weigh our intentions up in consideration when making decisions, because rational constraints do not come in degrees - they cannot be assigned deliberative weight. In Bratman's words: 'intentions do not provide reasons for action to be placed on the scale with desire-belief reasons in determining what to do' (1987, p.34). Clearly, intentions play a crucial role in decision-making, but we do not make judgements about our intentions in the same way that we do with those things that are apt for taking as reasons. Imagine this disagreement characterised by the intentions that these decisions function to bring about. If Bob has an intention that he and Bill see family, and Bill has an intention that they see friends, there is intentional inconsistency. It is easy to see how the decisions that function to bring these intentions about in the other disagree. Bob and Bill don't necessarily need to settle their family vs. friends disagreement by making a full decision about who to visit. Either of these decisions would rationally constrain them to just one of these options, in virtue of the associated intention, but this was not the intended outcome of this preliminary disagreement. They just needed to agree on how to weigh up these options, on balance, with all of their other reasons. There are likely to be many reasons weighing in favour of each trip destination, and while it would be fruitful to come to an agreement about matters such as the importance of who they visit, they don't need to be totally rationally constrained by this agreement.

Notice a consequence of this view. It is possible to have at least two kinds of normative disagreement. The first kind concerns the kind of commitments that put rational constraints on our plans. The second kind concerns the deliberative weight we assign when we make judgements about reasons. I want to suggest one way that the expressivist can take this distinction, although I do not endorse it as a consequence of this account. This distinction could mark a difference between deontic and evaluative disagreements. If this were so, deontic disagreements are disagreements in decision, evaluative disagreements are disagreements in reason-judgements. Deontic judgements, then, would happen at the level of decisions. The expressivist could say something like: decisions are 'ought' judgements. This would mean that judgements about permissibility are something like the negation of decisions. If I say, for example, that murder is impermissible, I seem to be making a deontic claim. That is, I am saying that it is not to be done. This can plausibly be represented at a decision on behalf of everybody to not murder. If I were to say that it were permissible, this could be the (external) negation of the decision to not murder. Evaluative judgements, on the other hand, would happen at the level of reason-judgements. If I say that murder is bad, this can be seen as an evaluative judgement, assigning deliberative weight to not-murdering. Ultimately, the evaluative judgement functions as the basis on which to decide, but its specific function is to weigh in favour of not murdering. This is one possible way to understand this consequence of the view, and the expressivist might wish to reject it, or does not want the burden of the extra commitments it entails. I do not endorse the evaluative/deontic distinction as a component of the view that I am defending, but this is a possibility that might be worth exploring for the expressivist.

I have shown how the theory of disagreement that I am endorsing seems to give rise to at least two kinds of normative disagreement. I have argued that these each could resemble deontic or evaluative normative disagreement. If there is a reason to believe in the existence of a real difference between these kinds of judgements, then this account of normative disagreement has the advantage of being able to

naturally draw the distinction. The claim that this does constitute the difference between normative and evaluative disagreement is substantial, and exploring it properly would be outside of the scope of this essay. Nonetheless, the two levels of disagreement, one concerning rational constraints on decisions and intentions, one concerning reasons for making decisions, is an interesting consequence of the view that I am endorsing. For one thing, it is a natural extension of the theories that I am taking up in explaining disagreement. It meets the objection because it explains disagreement between normative attitudes that are not fully-fledged decisions or judgements, like certain affective attitudes. In any case, it tells us that normative disagreement doesn't only happen when judgements would lead to intentional inconsistency.

3.3.5 Objection: There is More Than One Normative Attitude

Another objection to the above account might be that normative disagreement doesn't just pertain to what is to be done, or whether something counts as a reason for acting simpliciter. There are all sorts of types of normative disagreements, and as such, characterising disagreement in the way presented above oversimplifies the matter. How is, for example, moral disagreement distinct from other types of disagreement? If the answer is that it is one distinct type of judgement about what to do *morally*, then is the supporter of this view committed to the existence of many oughts, one for every normative consideration? The worry here is that Simple Plan Expressivism concerns all-things-considered oughts, and it is not obvious how it can account for oughts of specific types, like the moral one. If it can account for different oughts, then, does there need to be a way to adjudicate between these oughts? If we have different types of normative judgements, presumably they can sometimes disagree. What somebody ought morally to do might well be different to what they self-interestedly ought to do. Do we need an account of how we can settle disagreements between these oughts? Do we need to insert an independent normative standard, such as what is referred to as 'ought simpliciter' to determine our all-things-considered judgements in light of these disagreements? I have already argued that intentions, thus decisions, are not things that can be taken as reasons and weighed up in consideration. Therefore, the expressivist needs an account of the difference between all-things-considered judgements and specific normative judgements, and to minimise controversial commitments, it would be best if the account doesn't appeal to something like ought simpliciter to resolve disagreement between distinct normative decisions.

3.3.6 Reply: 'Focussed' and All-Things-Considered Normative Judgements

In response to the objection above, I will now explore a question that is potentially pertinent to the normative expressivist. How are different types of normative disagreements to be distinguished? And, what happens when different normative attitudes give us conflicting 'oughts'? The account of normative disagreement that I have defended in this essay is grounded in Ayars' taxonomy of normative judgements. Ayars is talking about all-things-considered judgements about what to do. We need to also provide an account of how we can make judgements about specific normative attitudes, too. I will refer to these as 'focussed' normative judgements for simplicity. These 'focussed' judgements are ones that are made in line with a specific normative consideration. They could be moral, aesthetic, prudential, comedic, and so

on. One particular issue that the metaethical expressivist might be interested in is what makes moral judgements and disagreement distinct from these other kinds of judgements.

Of course, non-cognitivists about all-things-considered ought judgements can be cognitivists about some, or all, 'focussed' ought judgements. For example, it is possible to subscribe to the view presented so far, while maintaining that there is a fact of the matter about aesthetic value, and thus about what we ought to do aesthetically speaking. The theory of normative disagreement with the most explanatory power for expressivists, however, is likely to be the one that holds that all normative judgements are non-cognitive. As such, I will try to account for the objection in non-cognitivist terms.

The expressivist can argue that the difference between different types of normative judgements, take moral and aesthetic ones, is that moral judgements take into consideration moral reasons, aesthetic judgements take into consideration aesthetic reasons, and so on. How is this so? Reason judgements, recall, consist in weighing something in favour of action. Decisions are defined by their function of and guiding intentions based on reason-judgements. Moral judgements are those decisions that are made on the basis of specifically moral reasons. Moral disagreement, therefore, is normative disagreement about those decisions (or, for that matter, reason-judgements) that are made on the basis of reasons that are distinctly moral. Of course, the expressivist needs to account for the difference between these types of judgements about reasons. This is a psychological difference. The expressivist can say that some of our attitudes are moral ones, some are aesthetic ones, etc., and that this difference is an empirical fact about our psychology. Presumably, somebody who subscribes to the view that 'focussed' normative judgements are non-cognitive has such an account of the psychological difference between different types of reasons-judgements in mind already. One option is to offer an explanation as to why we have different types of desires and dispositions that can give different deliberative weight to the same content. If the expressivist can explain what it is to desire *x* for moral reasons, but *not-x* for aesthetic reasons, without appealing to the content of *x*, then they can use this to explain the difference in types of reason-judgements. So, when those attitudes that are distinctively moral are treated as weighing in favour, the decision is a moral one. Moral decisions are made in light of moral reasons-judgements. Aesthetic decisions are made in light of aesthetic reasons-judgements, and so on.

This definition of 'focussed' decisions needs qualification, otherwise it might misconstrue the definition of decision that I adopted from the Simple Plan Expressivist view, though. Decisions that aren't all-things-considered judgements are not strictly answers to 'the practical question of what to do'. It seems as though they answer the question of what to do from a particular normative perspective. Perhaps they can be characterised as contingency decisions. For example, the judgement that 'aesthetically speaking, I ought to buy a new jacket' does not, by itself, guide intention (assuming that there exist reasons pertaining to whether to buy the jacket that are not aesthetic ones). It guides intention *iff* I am concerned with only aesthetics. They are conditional decisions. I can, for example, decide that aesthetically, I ought to buy the jacket, while fiscally, I ought not to. This combination of normative attitudes do not put me in disagreement with myself.

How should the expressivist explain these conditional decisions? As will become obvious, the answer draws on what Gibbard says about contingency planning (2003, p. 49), but is explained in terms of decisions. Under the Simple Plan Expressivist theory of normative judgement, I can make decisions for others. Presumably, as a natural extension of this, I can also make decisions for myself, that apply in circumstances slightly different to the one I find myself in currently. One straightforward example would be deciding for my future self. I can make a decision that functions to form the intention in me, tomorrow, to write 500 words. I don't necessarily need to form the intention now, but when the time comes, if all

goes well, I will intend to write 500 words. So, how does this help to explain conditional 'focussed' judgements? To use an intuitive example, suppose that I am sitting on a train, approaching my stop. I do not know on which side of the train the platform will be, thus, I form two conditional decisions. On one, I get up and exit the train on the left. On the other, I get up and exit the train on the right. If these decisions function to form intentions in the same agent, they would disagree. However, they do not. I make the former for versions of my future self where the platform is on the left, I make the latter for versions of my future self where the platform is on the right. To return to the jacket example, I can decide that in circumstances where I do not care about any considerations except aesthetic ones, I buy it, and that in circumstances where I do not care about anything but money, I do not buy it. The decisions function to form intentions in an agent who is distinct from myself currently, thus, it is a contingency decision. Importantly, the subjects of each decision are also distinct from each other. As such, the decisions do not function to form inconsistent intentions in the same agent, and do not disagree. The relevant agent in each scenario is a different person, and so the decisions do not preclude each others' functioning.

But what if we have the same normative considerations in mind? Surely, then we can disagree. Disagreements over what one ought to do aesthetically, for example, do rely on a hypothetical scenario where the antecedent is met. Imagine that I am disagreeing with you over whether I should buy the jacket. I might say 'if money were no object, I should buy the jacket', and you might reply 'even if you had all of the money in the world, you should not buy it'. These are not conflicting decisions about what I am to do, rather, they are decisions on behalf of a version of me for which money is not a consideration. We are plausibly deciding for the same agent, because our decisions each pertain to a version of me who does not need to worry about the cost of things. There is clearly a sense in which they preclude each other's functioning, even if this is conditional on there being a hypothetical version of me such that these decisions apply to. Given that normative judgements are decisions that function to settle practical questions, it makes sense that in these abstract 'focussed' disagreements, we still aim to settle the question of what a particular agent is to do, even if that agent does not exist. Normative judgements are still practical attitudes, and normative disagreements are still practical disagreements. The account on offer can explain this.

4. Conclusion

In this essay, I have argued that a satisfactory account of normative disagreement is available to normative expressivists. I have done so by appealing to Neil Sinclair's general account of disagreement in thought, and Alisabeth Ayars' theory of normative judgement. The core thesis is that disagreement in normative judgement can be explained in terms of the preclusion of a set of judgements jointly bringing about an intention in an agent. I have argued that this theory can meet the several success criteria outlined in this essay's survey of existing accounts. The theory of disagreement is unified because it appeals to Sinclair's general account of disagreement in thought as the preclusion of mental states' joint functioning. The theory can capture agent-relative judgements because it appeals to Ayars' theory of normative judgements as decisions. The appeal to Ayars' theory also explains the symmetry between our judgements for others and our own contingency plans. The theory I defend avoids the worry that the rational requirement for corealiseability overgenerates disagreement, which was an objection to Ayars' account. Furthermore, if the expressivist is sympathetic to the normative theory of disagreement, this theory can explain how each of the conditions of Bex-Preistley and Shemmer's analysis is met.

Upon examining potential objections to this view, several interesting consequences were discovered. Firstly, it is compatible with a Bratmanian view of intentions. This means that it can explain why judgements can be asserted individually, but not together, or vice-versa. This also explains how our decisions might implicitly function to form intentions in agents other than the subject of the decision. Secondly, the view can explain normative disagreement at the level of reasons. This type of disagreement does not concern which rational constraints to put on an agent's intentions, but how to weigh beliefs and desires in or out of favour of certain decisions. Thirdly, the account can distinguish between types of normative judgement, given that the expressivist has an account of the difference between normative attitudes. It can explain disagreement in 'focussed' normative judgements in terms of conditional decisions, retaining the practical nature of normative judgements.

The account defended in this essay understands normative disagreement in terms of the states of mind that normative judgements function to bring about. There are likely many ways that the expressivist can take up this general approach. The one that is defended here meets criteria that previous accounts have failed to meet, and has interesting consequences that open up avenues for further inquiry. The expressivist may use it to pursue a normative theory of disagreement in the way suggested. They may explore whether it can account for the difference between evaluative and deontic judgements. Alternatively, they might be able to offer a more fleshed out account of particular 'focussed' normative judgements. Whether or not these pursuits would be fruitful, the account on offer can make sense of normative disagreement on wholly expressivist terms.

References

Ayars, A. 2022. Deciding for Others: An Expressivist Theory of Normative Judgment. Philosophy and Phenomenological Research 105 (1). pp. 42-61

Ayars, A. An Explanation of the Essential Publicity of Practical Reasons. In: Shafer-Landau, R. (ed.). Oxford Studies in Metaethics, Volume 19. Online edn. Oxford University Press. pp. 80-104.

Bex-Priestley, G. and Shemmer, Y. 2017. A normative theory of disagreement. Journal of the American Philosophical Association, 3 (2). pp. 189-208

Blackburn, S. 1998. Ruling passions: a theory of practical reasoning. Oxford: Oxford University Press

Bratman, M. 1987. Intention, Plans, and Practical Reason. Cambridge: Cambridge, MA: Harvard University Press

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

Dreier, J. 2006. Negation for Expressivists: A Collection of Problems with a Suggestion for their Solution. Oxford Studies in Metaethics 1:217-233

Estlund, David (2020). Utopophobia: On the Limits (If Any) of Political Philosophy. Princeton: Princeton University Press

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

Köhler, S. 2022. Practical Expressivism. Philosophical Review 131 (4):515-518

Lennertz B. 2021. Noncognitivism and the Frege-Geach Problem in Formal Epistemology *Philos Phenomenol Res*.; 102: 187–211

MacFarlane, J. 2014. Assessment Sensitivity: Relative Truth and its Applications. Online edn. Oxford: Oxford University Press

Mellor, R. and Shea, M. 2025. What Are We to Do? Making Sense of 'Joint Ought' Talk. Philosophical Studies 182 (3):705-724

Moon, A. and Jackson, E. 2020. Credence: A Belief-First Approach. Canadian Journal of Philosophy 50 (5):652–669

Núñez, C. forthcoming. Expressivism, Moral Psychology and Direction of Fit. In: Copp, D. and Rosati, C. (eds.) The Oxford Handbook of Metaethics

Ridge, M. 2012. Disagreement. Philosophy and Phenomenological Research 86 (1). pp. 41-63

Ridge, M. 2014. Impassioned Belief. New York, NY: Oxford University Press

Shea, M. Why Plan-Expressivists Can't Pick Up the Moral Slack. In: Shafer-Landau, R. (ed.). Oxford Studies in Metaethics, Volume 19. Online edn. Oxford University Press. pp. 129-151

Shemmer, Y. and Bex-Priestley, G. 2021. Disagreement without belief. Metaphilosophy 52 (3-4):494-507

Sinclair, N. 2021. Practical Expressivism. online edn. Oxford: Oxford University Press

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

Toppinen, T. 2022. Neil Sinclair, Practical Expressivism: A Metaethical Theory. Ethics 133 (1):152-156

Worsnip, A. 2019. Disagreement as Interpersonal Incoherence. Res Philosophica 96 (2). pp. 245-268

Zeman, D. 2020. Minimal Disagreement. Philosophia 48 (4):1649-1670