

Reversibility and Disagreement¹

(this paper supersedes our earlier draft manuscript, ‘reflections on epistemic modals’)

Abstract

Several philosophers have recently argued that a number of expressions that have traditionally been regarded as context-sensitive (including epistemic modals, probability operators, conditionals, and predicates of personal taste) in fact have semantic value that are invariant across contexts of utterance. The main arguments for this invariantist position have relied on the following observation: speakers who make assertions involving these expressions seem to *disagree* with one another, even when their assertions are made in very different contexts. We show, however, that while the phenomenon of disagreement may appear to support invariantism, another phenomenon, which we call *reversibility*, raises serious problems for this position. Each of the expressions in question can be used to construct sentences that one can rationally and sincerely assert even while knowing that in the future one will rationally and sincerely assert their negations. We argue that such reversibility calls into question the claims about disagreement that invariantism is meant to explain, as well as the kind of explanation of disagreement that invariantism allows.

Contextualists, about a given type of expression, claim that the contribution that expression makes to the content or truth value of an assertion depends on the context in which this assertion is made, whereas invariantists deny this. Recently, arguments have been given for favoring invariantist accounts over more standard, contextualist accounts of a number of expressions, including predicates of personal taste (e.g., ‘tasty,’ ‘dreamy’),² epistemic modals

¹ Many thanks to Stephen Finlay, John MacFarlane, Scott Soames, Julia Staffel, Michael Titelbaum, and all of the others who have given us helpful feedback on this project.

² Lasersohn [2005], Stephenson [2007a].

(‘might,’ ‘must’),³ deontic modals (‘ought,’ ‘may’),⁴ probability operators (‘probably,’ ‘certainly’),⁵ and indicative conditionals,⁶ among others.⁷

Invariantists have argued that their view is preferable to contextualism because only their view accounts for disagreement and related phenomena. For an invariantist theory seems to predict, correctly, that Popeye and Bluto disagree when Popeye says ‘spinach is tasty’ and Bluto says ‘spinach is not tasty,’ whereas a contextualist theory, according to which each one is asserting something about his own taste, seems to predict that they do not disagree. More generally, if we define a *contested sentence* as a sentence containing an expression of one of the kinds listed above (predicates of personal taste, epistemic modals, etc.) but no other context-sensitive expressions, invariantism seems to entail that the following is true, whereas contextualism seems to entail that it is false:

disagreement thesis: For any contested sentence, *s*, and competent speakers, *x* and *y*, if *x* sincerely assertively utters *s* in one context, and *y* sincerely assertively utters $\sim s$ in another context, *x* and *y* thereby disagree.⁸

While the phenomenon of disagreement may seem to favor invariantism over contextualism, there is another phenomenon that seems to favor contextualism over invariantism: namely, that it is often perfectly rational to sincerely utter a contested sentence while believing that in the future one will sincerely utter its negation. Thus, the fourteen-year-old Audrey could rationally and sincerely utter ‘high school football players are dreamy,’ while expecting that when she is in college, she will sincerely utter ‘high school football players are *not* dreamy’. Even clearer cases of such rational anticipated reversals can be found in the other contested domains, as we will show later. Thus, in general,

³ Egan, *et al.* [2004], MacFarlane [forthcoming], Stephenson [2007a].

⁴ Kolodny and MacFarlane [unpublished a].

⁵ Egan [2007], MacFarlane [forthcoming].

⁶ Weatherson [2009], Stephenson [2007b], Kolodny and MacFarlane [unpublished b].

⁷ Several of our cases – epistemic modals, probability operators, indicative conditionals, and deontic modals – fall under the heading of being ‘evidence relative.’ Predicates of personal taste constitute a different but also important category that many have held to exhibit similar behavior. Note that in grouping these expressions together as ‘contested expressions,’ we do not mean to be taking a stand on whether the same treatment is appropriate for each domain – only noting that a similar dialectic arises for each. Similar dialectics have also arisen for other topics, including gradable adjectives (Cappelen and Lepore [2004], Richard [2004]) and ‘knows’ (Cappelen and Lepore [2004], Stanley [2005]).

⁸ For example – just to focus on two of the most prominent of this paper’s targets – it is clear from the extended discussion of disagreement in MacFarlane [unpublished] as well as in MacFarlane [2007] that he is committed to the disagreement thesis, and the disagreement thesis is presupposed throughout Egan [2010].

reversibility thesis. In each of the contested domains, there are contested sentences that a fully rational speaker could sincerely assertively utter, under standard conditions,⁹ while believing that later she will sincerely assertively utter their negations.

We will be arguing that the invariantist cannot reconcile the reversibility thesis with the disagreement thesis, and that this fact undermines the plausibility of the invariantist's position.

Our argument will proceed as follows. We will begin, in part 1, by setting out the various claims under dispute. In section 1.1, we clarify the distinction between contextualism and invariantism, as well as the separate distinction between relativism and non-relativism. In section 1.2, we discuss the disagreement thesis, and its role in the arguments for invariantism. And in 1.3, we argue for the reversibility thesis by providing instances of reversibility in each of the contested domains. In part 2, we argue that there is a *prima facie* incompatibility between invariantism and the reversibility thesis, which can be reconciled only if the invariantist adopts a relativist version of invariantism.¹⁰ In part 3, we argue that there is a *prima facie* incompatibility between relativist invariantism and the disagreement thesis, and that none of the ways in which the invariantist might attempt to solve this problem can succeed, without sacrificing reversibility once more. We conclude, in part 4, by offering a diagnosis of the invariantist's inability to reconcile reversibility and disagreement. And in light of this diagnosis, we argue that we should reject the disagreement thesis, and with it the main arguments for favoring invariantism over contextualism.

1 Stage-Setting

1.1 A Taxonomy of Views

Consider the sentence 'water might be an element.' There are several views one might take about this sentence. On a *propositional* view, its semantic contribution is a proposition, understood as something which determines truth-conditions for the sentence. On a *non-*

⁹ The relevant sense of 'standard conditions' will be explained in section 1.3. Roughly, standard conditions are those in which the agent expects to remain fully rational and to undergo no information loss.

¹⁰ Note that, as we have defined invariantism, it is compatible with relativism. We defend this usage in section 1.1.

propositional view, such as expressivism, this is not the case. In this paper, we will be concerned only with propositional views.¹¹

Propositional views come in many varieties. Contextualist views affirm, while invariantist views deny, that the truth of an assertion depends on the context in which it is made.¹² On most versions of contextualism, the reason the truth of an assertion depends on the context of utterance is that *the proposition asserted* depends on this context. Alternately, one could hold that the proposition asserted by an assertive utterance of a sentence is the same across contexts, but that the truth value of such an assertion depends on the context in which it is made.¹³

Another exhaustive distinction among propositional views is the distinction between relativism and non-relativism. Relativists assert, while non-relativists deny, that the truth value of a contested sentence depends on the context *from which this assertion is evaluated*. According to most relativists, there is a single, context-invariant, fact about what proposition is asserted by a given utterance of such a sentence, but whether this proposition counts as true depends on the context from which it is assessed. However, on some relativist views, the reason why the truth of an assertion depends on the context from which it is assessed is that *the proposition asserted* varies with the context of assessment.¹⁴ Nothing in what follows will turn on the difference between these two forms of relativism.

Note that, as we are defining these terms, the contextualist/invariantist distinction is orthogonal to the relativist/non-relativist distinction. Hence, one can be a relativist contextualist, and hold that the truth value of an assertion depends both on the context in which it is made and on the context from which it is assessed. Indeed, it will emerge from the argument of the final section that if any relativist view is plausible, it is relativist contextualism.

1.2 The Disagreement Thesis

¹¹ Expressivist views are most familiar in the moral case, but may also be defended for epistemic modals (Price [1983], Yalcin [unpublished]), conditionals (Gibbard [1981]), knowledge (Chrisman [2007]), and other topics. See particularly Gibbard [1990] and Schroeder [2008]. Another significant class of non-propositional theories is that of *dynamic* theories, for which see particularly Veltman [1996] and Gillies [2004].

¹² Note that this formulation counts invariant ‘propositional radical’ views like that in Bach [forthcoming] as contextualist. This is as it should be – the debate within semantic theory about how much of context-dependence needs to follow from fully elaborated semantic *characters* is largely orthogonal to the debate over whether there is context dependence in the truth values of assertions in the first place.

¹³ This latter view related to, but not exactly the same as, the view that MacFarlane [2009] calls ‘nonindexical contextualism.’

¹⁴ See Egan, *et al.* [2005].

Disagreement has long been central to the debate between contextualists and invariantists.¹⁵ On the simplest version of contextualism about predicates of personal taste, when Popeye says ‘spinach is tasty,’ his assertion is true just in case spinach tastes good *to Popeye*, while when Bluto says ‘spinach is not tasty,’ his assertion is true just in case spinach does not taste good *to Bluto*. But since it could be true both that spinach tastes good to Popeye and that it doesn’t taste good to Bluto, this theory seems to predict that Bluto and Popeye do not disagree in making these claims. Similarly, on the simplest contextualist theory of epistemic modals, when Thales says ‘water might be an element,’ his assertion is true just in case water’s being an element is compatible with what is known in *his* context of utterance, while, when Cavendish says ‘it is not the case that water might be an element,’ his assertion is true just in case water’s being an element is incompatible with what is known in *his* context of utterance. And since water’s being an element could be compatible with what is known in one context and incompatible with what is known in the other, this view seems to imply that Thales and Cavendish don’t disagree.

More sophisticated versions of contextualism might avoid the conclusion that there is no disagreement in these particular cases. But *any* contextualist theory will predict that there are *some* cases in which one speaker assertively utters one contested sentence in one context, and another speaker assertively utters its negation in another context, and the two assertions could both be true. Invariantism, however, implies the opposite. The invariantist can therefore explain why the two speakers disagree in such cases, so long as she assumes that people disagree whenever they sincerely make assertions that cannot both be true. Given this assumption, invariantism entails the disagreement thesis, and hence it explains disagreement in many contexts in which contextualism seems to predict that there is none.

It has also been argued that invariantism does better than contextualism in relation to truth and falsity ascriptions, and in relation to retraction. Suppose Thales says ‘water might be an element,’ and subsequently learns that water is a compound. Invariantists argue that, after learning this, Thales should say ‘I take it back, what I said was false—water can’t be an element since it’s a compound.’ And invariantists can explain why this would be so: for Thales is now in a position to make a contrary assertion whose truth is incompatible with that of his original assertion, and so he should now regard this original assertion as false. Contextualism,

¹⁵ This dialectic isn’t new; compare the criticisms of speaker subjectivism in ethics in Stevenson [1937].

however, seems to predict the opposite, since, on this view, given the context in which Thales' original assertion was made, it is true, and so it would seem that he should stand by it.¹⁶

Since the problem of disagreement has been central to the debate between contextualists and invariantists, it will be our main focus. However, in the concluding sections we will argue that some of the problems we raise for the invariantist account of disagreement extend to the invariantist accounts of the closely related phenomena of retraction and truth and falsity ascription.

1.3 The Reversibility Thesis

Let us say that a sentence is *reversible* just in case it instantiates the reversibility thesis, that is, just in case a fully rational speaker could sincerely utter this sentence, under *standard conditions*, while believing that she will later sincerely utter its negation. The reversibility thesis states that there are reversible sentences in each of the contested domains. The reference to 'standard conditions' in our formulation of the reversibility thesis is important. For if we remove this qualification then just about any sentence will count as reversible. That is, for just about any sentence that one can rationally believe, there are *non-standard conditions* under which one could rationally assert this sentence while believing that later (say, after being brainwashed, or after becoming confused about the meanings of one's words, or after undergoing some other non-rational change in one's beliefs) one will sincerely utter the negation of this sentence. What is special about the contested expressions is that they give rise to the possibility of such reversals under conditions in which none of these distorting factors is present, i.e., in which one expects to remain fully rational, in which one expects to lose no information, and in which one understands, and expects to continue to understand, the meanings of one's words. The reversibility thesis states that under these *standard conditions*, a fully rational agent can sincerely utter sentences, in each of the contested domains, while believing that she will later sincerely utter their negations.

We have already noted a reversible sentence involving a predicate of personal taste: a fully rational young Audrey could sincerely say 'high school football players are dreamy' while

¹⁶ For more on retraction, see MacFarlane [2007]. On truth-ascriptions see Cappelen and Lepore [2004], Egan, *et al.* [2005], and MacFarlane [2007]. On the closely related phenomenon of denial, see Kölbel [2002], chapter 3 and Stephenson [2007a]. For contextualist replies to these arguments, see von Fintel and Gillies [2008] and Björnson and Finlay [manuscript].

believing she will later sincerely utter the negation of this sentence. As an illustration involving epistemic modals, consider the following case:

Old News: Rita is writing a story for the *Morning Bugle* on a local murder investigation. Since, that morning, it is unknown whether Axeworthy is the murderer, Rita asserts sentence m : ‘Axeworthy might be, and might not be, the murderer.’ She knows, however, that in the afternoon the DNA test will be completed and its results announced, establishing whether Axeworthy is the murderer. Rita knows that the writers for the *Evening Standard* are very sloppy, and she predicts that they will include sentence m in this evening’s edition. She expects that when she reads this, she will exclaim: ‘Nonsense! It is not the case that Axeworthy might be, and might not be, the murderer.’

Once again, Rita appears to be perfectly rational in sincerely asserting m while believing that she will later sincerely assert $\sim m$.

The same goes for probability operators. Before the winning lottery ticket has been revealed, Winona could rationally and sincerely utter l (‘each of the lottery tickets probably is not the winning ticket,’) while believing that, after the winning ticket has been revealed, she will sincerely utter $\sim l$. The same goes for the normative ‘ought,’ as the following case illustrates.

Three Envelopes: Mable and Mabos are watching a movie in which the protagonist, Chester, must choose among three envelopes. Chester is entirely ignorant of their contents, and is about to choose among them at random. Mable and Mabos, however, know that the first envelope contains \$900, and that one of the remaining two envelopes contains \$1000, while the other is empty. And so Mabos says ‘Chester ought to take the first envelope. For it’s sure to contain \$900.’ He knows, however, that later in the film it will be revealed which envelope contains the \$1000. And he also knows that he and Mable will rewatch the film. And so he knows that if the \$1000 turns out to be the second envelope, then when they rewatch the film he will say: ‘It is not the case that Chester ought to choose the first envelope. For he ought to choose the second.’ And if the \$1000 turns out to be in the third envelope, he will say: ‘It is not the case that Chester ought to choose the first envelope. For he ought to choose the third.’¹⁷

Once again, Mabos appears to be perfectly rational, in spite of this expected reversal in the sentences he sincerely utters.

Finally, there are reversible sentences involving indicative conditionals, as this next case illustrates.

¹⁷ This example is from Ross [2006]. It is structurally analogous to the mine shaft problem introduced in Regan [1980], and discussed in Parfit [forthcoming] and in Kolodny and MacFarlane [unpublished a]

Holmes's Reversal: Moriarty, Milverton, and Moran are the three possible suspects in a murder that was committed with an air-rifle. Believing that only one of the suspects had an air-rifle at the time of the murder, Watson says: 'Now all we need to find out is who had an air-rifle. For each of the suspects, if he had an air rifle, then he is the murderer.' Holmes, however, has deduced that in fact *two* of the suspects had air-rifles. And so he says: 'My dear Watson, it is not the case that if Moriarty had an air rifle, then he is the murderer, and it is not the case that if Milverton had an air rifle, then he is the murderer, and it is likewise not the case that if Moran had an air rifle, then he is the murderer.' (Call this conjunctive sentence r .) Holmes also knows that he will soon learn the identity of the innocent air-rifle owner, and that he will subsequently know, of each of the remaining two suspects, that if he had the air rifle, then he is the murderer. And so Holmes predicts that at this later time, he will assertively utter $\sim r$.

Thus, while not every contested sentence is reversible (e.g., 'water might be an element' is not reversible), we have shown that in each of the contested domains, there are reversible sentences, as the reversibility thesis states. This fact creates problems for the invariantist.

2 The Problem of Reversibility

In order to show that reversibility creates problems for the invariantist, we argue, in section 2.1, that invariantism and the reversibility thesis are jointly incompatible with the principle of reflection. Consequently, if the invariantist wants to endorse the reversibility thesis, she must maintain that the propositions we assert in uttering the contested sentences are not governed by the reflection principle. In section 2.2, we introduce one possible conception of these propositions that would explain why they are not governed by reflection, namely Andy Egan's relativist conception. And then we argue, in section 2.3, that *only* on a relativist conception of these propositions is invariantism compatible with reversibility. And so there is strong pressure on the invariantist to accept relativism. Such a move, however, will set the stage for a further dilemma confronting the invariantist's efforts to reconcile reversibility and disagreement, as we will show in part 3.

2.1 The Prima Facie Incompatibility between Invariantism and Reversibility

The reversibility thesis, conjoined with invariantism, appears, at least prima facie, to be incompatible with the principle of reflection. To see why this is so, assume that the reversibility thesis is true, and hence that a fully rational individual, whom we may call 'Rachel,'

sincerely utters a contested sentence, s , while believing that at a future time, t , she will sincerely utter $\sim s$. Assume, further, that invariantism is true, and hence that speakers who assert s assert the same proposition in any context; let p be this proposition. Since Rachel now sincerely asserts s , and understands the meanings of her words (since we are assuming standard conditions obtain), it follows that she must now believe p . And since we are assuming that assertive utterances of s assert p in any context, it follows that if Rachel were to assertively utter $\sim s$ at t , she would thereby assert $\sim p$. And since Rachel is fully rational, understands the meaning of her words, and expects to continue understanding the meaning of her words (again, because standard conditions obtain), it follows that she must believe that when, at t , she sincerely utters $\sim s$, she will thereby sincerely assert $\sim p$, while understanding the meaning of her words, and hence that she must then believe $\sim p$. Thus, from the assumption that the reversibility thesis and invariantism are both true, we have deduced that, under standard conditions, a fully rational individual, Rachel, can believe a proposition, p , while believing that she will later believe its negation.

But this is incompatible with even a very weak formulation of the principle of reflection. On its standard formulation, this principle states the following.

reflection: If one is fully rational, then for any proposition, h , any future time, t , and any level of credence, x , that one thinks one may have in h at t , one's current credence in h , conditional on the supposition that at t one's credence in h will be x , must itself be x .¹⁸

Assuming that a fully rational individual will recognize that a necessary condition for rationally believing a proposition is having a reasonably high credence in this proposition (anything greater than $(\sqrt{5}-1)/2$, or approximately .618, will do), the principle of reflection straightforwardly entails that one cannot rationally believe a proposition while believing that in the future one will believe its negation.¹⁹

¹⁸ See van Fraassen [1984].

¹⁹ Proof: Suppose a fully rational individual, Rachel, believes p while believing that at future time t she will rationally believe $\sim p$. Assume that to rationally believe a proposition one's credence in it must exceed $(\sqrt{5}-1)/2$, and that this fact will be recognized by any fully rational individual. It follows that Rachel now has credence greater than $(\sqrt{5}-1)/2$ in p , and that she has credence greater than $(\sqrt{5}-1)/2$ that at t she will have credence greater than $(\sqrt{5}-1)/2$ in $\sim p$. Hence, by the principle of reflection, together with the total probability theorem, it follows that her current credence in $\sim p$ must be greater than $((\sqrt{5}-1)/2)^2 = (3-\sqrt{5})/2$, and hence that her current credence in p must be less than $1 - (3-\sqrt{5})/2 = (\sqrt{5}-1)/2$, which contradicts our hypothesis.

Of course, it is well known that one can rationally violate the principle of reflection in cases in which one expects to be irrational or to lose information.²⁰ But even if we restrict the principle of reflection to exclude such cases, it will still imply that, under *standard conditions*, in the sense we defined earlier, one cannot rationally believe a proposition while believing that one will later believe its negation. And yet, as we have just seen, the invariantist is committed to claiming that if the reversibility thesis is true, then such violations of reflection under standard conditions can be rational. Thus, the conjunction of invariantism and the reversibility thesis is incompatible with even a highly restricted version of the reflection principle.

2.2 Egan's Relativist Solution

One way for the invariantist to respond to this problem is to maintain that the propositions we assert in uttering the contested sentences are exempt from the principle of reflection, because this principle applies only to ordinary propositions, whereas the propositions in question are instead *de se* or *centered* propositions.

De se or centered propositions are intended to serve as the objects of self-locating beliefs.²¹ On the view we are considering, while ordinary propositions indicate what the world is like, and determine a set of possible worlds, centered propositions indicate what one's location in the world (perhaps in addition to indicating what the world is like). They indicate, in particular, *who* one is and *when* the present time is. Thus, rather than simply determining a set of possible worlds, a centered proposition determines a set of *centered words*, where a centered world is a triple, $\langle w, t, i \rangle$, consisting in a possible world, w , a privileged time, t , and a privileged individual, i . To believe a centered proposition corresponding to a given set of centered worlds is to believe that one of the worlds in this set represents where, when and who one is. Thus, to believe the centered proposition associated with the set of centered worlds in which the privileged individual is Waldo is to believe that one is Waldo, while to believe the centered proposition associated with the set of centered worlds in which the privileged time is April 1st, 2011 is to believe that it is now April 1st, 2011. Let us label these centered propositions *I am Waldo* and *It is April 1st, 2011*, respectively.

Andy Egan has proposed that what sentences involving epistemic modals express are such centered or *de se* propositions. On his account,

²⁰ See Christensen [1991].

²¹ See Lewis [1979].

The semantic value of ‘Bond might be in Zurich’ [is] the set of all $\langle w, t, i \rangle$ triples such that it’s compatible with all the facts that are within i ’s epistemic reach at t in w that Bond is in Zurich ... So, believers in this proposition have a belief about their own epistemic circumstances: they take themselves to be in a situation in which none of the evidence that’s within their reach rules out Bond’s presence in Zurich.²²

Egan has offered similar views of other contested expressions, including predicates of personal taste.²³

Unlike ordinary propositions, centered propositions are not governed by the principle of reflection. One’s current credence in a centered proposition, conditional on the supposition that one’s credence in this proposition will later be x , can rationally differ radically from x , even under standard conditions. Suppose, for example, that on April 1st, 2011, Wendy knows that one day hence she will know that it is no longer April 1st, 2011. In this case, Wendy’s credence in the centered proposition *It is now April 1st, 2011*, conditional on the supposition that her credence in it tomorrow will zero, may rationally be much greater than zero, and could even be as great as one. Hence, she could fully believe this proposition (thereby locating herself at April 1st, 2011) while believing that tomorrow she will believe its negation (thereby locating herself within the range of times other than April 1st, 2011).

Similarly, in the morning, Rita could rationally locate herself in a situation in which the available evidence *does not* rule out either Axeworthy’s being the murderer or his not being the murderer, while believing that in the evening she will locate herself in a situation in which the available evidence *does* rule out one of these alternatives. But on Egan’s view, to believe the proposition one asserts in uttering sentence m (‘Axeworthy might be, and might not be, the murderer’) is simply to locate oneself in a situation of the former kind, and to believe the proposition one asserts in uttering $\sim m$ is simply to locate oneself in a situation of the latter kind. It follows that, on Egan’s view, Rita could rationally believe the former proposition while believing that in the evening she will believe the latter proposition. Thus, Egan’s view is compatible with the claim that, under standard conditions, Rita could rationally and sincerely utter m while believing that she will later sincerely utter $\sim m$. And so his view can accommodate reversibility.

Egan’s view is a form of *relativism*, as it implies that the assertions under consideration express propositions whose truth values can vary when evaluated from different contexts

²² Egan [2007, 8].

²³ See Egan [2006], [2011].

within the same world. For while ordinary propositions are assigned truth values relative to worlds, centered propositions are assigned truth values relative to centered worlds. Thus, where w is the actual world and t is the present time, *I am Waldo* is true relative to $\langle w, t, \text{Waldo} \rangle$ and false relative to $\langle w, t, \text{Wendy} \rangle$. Similarly, if Egan’s view is correct, then the proposition one asserts in uttering m will be true relative to $\langle w, \text{this morning}, \text{Rita} \rangle$ and false relative to $\langle w, \text{this evening}, \text{Rita} \rangle$.²⁴

What Egan’s view demonstrates is that the invariantist can reconcile her view with the reversibility thesis if she adopts relativism. In the next section, we will go further, and argue that the invariantist *must* adopt relativism if she is to endorse the reversibility thesis.

2.3 Why Solving the Problem of Reversibility Requires Relativism

Suppose non-relativist invariantism is true. Hence, where p is the proposition asserted in the utterance of a contested sentence, if an assertion of p , in a context, C_1 , is true as assessed from C_1 , then an assertion of p in any other context, C_2 , is true as assessed from C_2 . Presumably, the truth value of a belief that p , in a given context, is the same as the truth value of an assertion of p made in that context. Hence, it follows from non-relativist invariantism that if a belief that p , in context C_1 , is true as assessed from C_1 , then a belief that p in C_2 is true as assessed from C_2 .

Now consider, once more, the case in which, in the morning, Rita assertively utters m (‘Axeworthy might be, and might not be, the murderer’) while believing that in the evening she will assertively utter $\sim m$. As we saw in section 2.1, the invariantist view implies that there is a proposition, p_m , such that Rita rationally believes p_m in the morning while believing that in the evening she will rationally believe $\sim p_m$. Now, presumably, a rational individual will believe a proposition only if she has compelling evidence that this belief is true as assessed from the context she occupies. And so since, in the morning, Rita believes that in the evening she will *rationally* believe $\sim p_m$, she must believe that, in the evening, she will have compelling evidence for the truth of this belief, and hence for the falsity of the contrary belief that p_m , *as assessed from the context she occupies in the evening*. Consequently, assuming non-relativist invariantism, Rita

²⁴ Strictly speaking, this claim is compatible either with relativism or with a version of contextualism according to which the truth value, but not the content, of an assertion depends on its context of *utterance*, not its context of evaluation – the view MacFarlane calls ‘nonindexical contextualism.’ However, it is clear that Egan’s view is relativist rather than contextualist (see Egan [2007]). Further, the arguments we give in the next section apply as much to nonindexical contextualism as to relativism.

must likewise believe that, in the evening, she will have compelling evidence for the falsity, *as assessed from the context she occupies in the morning*, of the belief she has in p_m in the morning.

But since we are assuming standard conditions obtain, Rita must believe that between now and the evening, she will lose no information. Hence, Rita must now believe that she does not currently possess any evidence that defeats the compelling evidence she will later have for the falsity of her current belief as assessed from her current context. But it would hardly be rational for Rita now to believe p_m , when she believes that later, she will have compelling evidence that this present belief is false, and when she also believes that she now has no evidence to defeat this future evidence. And if, in the morning, Rita could not rationally believe p_m , then it follows that she could not rationally and sincerely assert p_m , and hence that she could not rationally and sincerely utter m . But we are assuming that this is precisely what Rita does. And so we have arrived at a contradiction.

The only plausible way to block this contradiction, while retaining both invariantism and the reversibility thesis, is to adopt relativism, and to maintain that the truth value of an assertion (and, correspondingly, of a belief) that p_m , as assessed from the context Rita occupies in the morning, can differ from its truth value as assessed from the context she occupies in the evening. On this relativist view, in the morning, Rita could rationally believe that her current belief that p_m is false as assessed from the context she will occupy in the evening, without believing that this belief is false relative to the context she now occupies. Hence, by adopting relativism, the invariantist can resist the conclusion that Rita cannot now rationally believe p_m .

3 The Problem of Disagreement

We have just argued that the invariantist must adopt relativism if she is to endorse the reversibility thesis. In what follows, we will argue that the adoption of relativism creates serious problems for any invariantist account of disagreement. We begin, in section 3.1, by arguing for a prima facie incompatibility between relativism and the disagreement thesis. In section 3.2, we consider John MacFarlane's assertion-based account of disagreement, and show how its pragmatic features allow it to account for disagreement even within a relativist framework. Then, in section 3.3, we show that the very features of MacFarlane's view that enable it to account for disagreement render it incompatible with the reversibility thesis. Thus,

while the relativist invariantist can have either disagreement or reversibility, she can't have both.

3.1 The Prima Facie Incompatibility Between Relativism and Disagreement

In order to see why relativist invariantism appears to be incompatible with the disagreement thesis, let's first consider the version of relativist invariantism defended by Egan. Recall that, on Egan's view, when a speaker asserts sentence e ('water might be an element') she asserts a centered proposition that locates her in a context in which water's being an element is compatible with all the available evidence. Thus, if Thales sincerely asserts e and Cavendish sincerely asserts $\sim e$, then Thales simply has a self-locating belief whereby he locates himself in one kind of context, and Cavendish simply has a self-locating belief whereby he locates himself in a different kind of context.

But we know independently that two people who locate themselves in different contexts needn't thereby disagree about anything at all. For example, if Thales locates himself in Greece, and Cavendish locates himself in England, they don't thereby disagree. No more, it would seem, should they disagree by virtue of the fact that Thales locates himself in a context where the available evidence is *compatible* with water's being an element, and Cavendish locates himself in a context where the available evidence is *incompatible* with water's being an element. Hence, on Egan's view, it would seem to follow that when Thales assertively utters e and Cavendish assertively utters $\sim e$, they could sincerely believe the propositions they assert without disagreeing. And so it would seem to follow, on Egan's view, that there needn't be any disagreement between Thales and Cavendish when they make these two assertions, contrary to the disagreement thesis. Hence, although Egan is committed to the disagreement thesis,²⁵ his theory appears, at least prima facie, to be incompatible with it.

Note that there are contexts in which it seems people *can* disagree in virtue of their self-locating beliefs. For example, if Eunice and Eustace are both riding unicycles, and each one takes herself to be located on the unicycle of Ustinov the Magnificent, then they would seem to disagree. For it presumably follows from their background beliefs that they can't both be on Ustinov's unicycle. Similarly, if Eunice and Eustace recognize that they are riding the same tandem bicycle, and Eunice locates herself in Yunnan while Eustace locates herself in

²⁵ See, e.g., Egan [2011], where he says that speakers like Thales and Cavendish are engaged in a 'genuine conflict.'

Uzbekistan, then again they seem thereby to disagree.²⁶ For it presumably follows from their background beliefs that it can't be both that Eunice is in Yunnan and that Eustace is in Uzbekistan. Thus, the problem for Egan's account is not that people can *never* disagree in virtue of having opposing self-locating beliefs, but rather that they cannot disagree in virtue of these self-locating beliefs when these beliefs, in conjunction with their relevant background beliefs, could all be true. To state this point as a general principle,

I'm OK, You're OK: For any two individuals, x and y , if x locates herself in a context with property A , and y locates herself in a context with property B , and it is consistent with the background beliefs of both that x is in an A context and y is in a B context, then x and y do not disagree, either implicitly or explicitly, in virtue of these differing self-locating beliefs.

Now it is presumably consistent with both Thales' and Cavendish's background beliefs that Thales is in a context in which water's being an element is compatible with the available evidence whereas Cavendish is in a context in which water's being an element is incompatible with the available evidence. Hence, it follows from the I'm OK, You're OK principle that if Thales locates himself in a context of the first kind, and Cavendish locates himself in a context of the second kind, they do not thereby disagree. And so this principle suffices to yield the conclusion that, on Egan's Theory, Thales and Cavendish can sincerely utter the negations of one another's epistemic modal sentences without disagreeing in their beliefs.

We can generalize this result to other relativist invariantist theories. For on any such theory, we should expect that the following situation could arise: For some contested sentence, s , expressing a proposition, p , and some pair of fully rational speakers, x and y , located in different contexts, x sincerely utters s , thereby asserting p , and y sincerely utters $\sim s$, thereby asserting $\sim p$, and it is consistent with both x 's and y 's background beliefs that p is true relative to x 's context and false relative to y 's context. (For example, s could be the sentence 'water might be an element', and x and y could be Thales and Cavendish, respectively.) We can now argue that, in such a situation, there needn't be any disagreement between the beliefs of x and y , despite their sincerely uttering the negations of one another's sentences. The argument proceeds as follows:

²⁶ Compare Egan [2007], where a similar example is used to illustrate a different point.

- (1) One can always rationally infer a proposition, p , from the centered proposition that p is true relative to the context one occupies. And, conversely, from any proposition, p , one can rationally infer the centered proposition that p is true relative to the context one occupies. E.g., if one believes *Snow is white* and *I am Waldo* one can rationally infer *Relative to my context it is true that snow is white* and *Relative to my context it is true that I am Waldo*. And conversely, if one believes the latter centered propositions, one can rationally infer *Snow is white* and *I am Waldo*.²⁷
- (2) Consequently, since x sincerely asserts p , and thus believes p , and since x is fully rational, x must believe the centered proposition, *p is true relative to my context*. And since y sincerely asserts $\sim p$, and thus believes $\sim p$, and since y is fully rational, y must believe the centered proposition, *$\sim p$ is true relative to my context*.
- (3) By I'm OK, You're OK, x and y do not disagree, either implicitly or explicitly, in virtue of the fact that x believes *p is true relative to my context* and y believes *$\sim p$ is true relative to my context*. I.e., they do not disagree in virtue of the fact that x locates himself in a context relative to which p is true, and y locates herself in a context relative to which $\sim p$ is true. For, by stipulation, it is consistent with their background beliefs that p is true relative to x 's context and false relative to y 's context.
- (4) But for any two speakers, i and j , and any centered or uncentered propositions, q , r , q' , and r' , if i and j do not disagree, either implicitly or explicitly, in virtue of believing q and r , respectively, and if i could rationally infer q' from q and j could rationally infer r' from r , then i and j do not disagree in virtue of believing q' and r' , respectively. (E.g., if Sophie and Helen do not disagree in virtue of the fact that Sophie believes that Socrates is wise and Helen believes that Socrates is Greek, then they can't disagree in virtue of the fact that Sophie believes that *someone* is wise and Helen believes that *someone* is Greek.)
- (5) Therefore, since, from (3), x and y do not disagree, either implicitly or explicitly, in virtue of believing *p is true relative to my context* and *$\sim p$ is true relative to my context*, respectively, and since, from (1), x could rationally infer p from *p is true relative to my context* and y could rationally infer $\sim p$ from *$\sim p$ is true relative to my context*, it follows from (4) that x and y do not disagree in virtue of believing p and $\sim p$, respectively.

²⁷ Note that Egan, *et al* [2005] are committed to (1). Also compare MacFarlane [2005, 333] and especially MacFarlane [manuscript], section 5 of chapter 5 on belief (most explicitly on page 139).

This argument shows that, given relativist invariantism, two speakers can sincerely assertively utter the negations of one another's sentences without disagreeing in virtue of their corresponding beliefs. This is very close to the conclusion that they can make these assertions without disagreeing, contrary to the disagreement thesis. But it leaves open one important possibility – that even though the speakers do not disagree simply in virtue of their beliefs, they *do* put themselves into disagreement with one another by their *assertions*. Let us call any account of disagreement according to which disagreement arises not from beliefs, but from the act of asserting, a *pragmatic* account of disagreement. What we have argued in this section is therefore that, at best, the only kind of disagreement for which relativists can allow, in the kind of case under consideration, is *pragmatic* disagreement. In the next section, we will consider one relativist invariantist theory on which disagreement has this pragmatic character, namely the theory offered by John MacFarlane.²⁸

3.2 MacFarlane's Pragmatic Account

MacFarlane comes close to acknowledging that, on his relativist view, when one speaker assertively utters a contested sentence and another speaker assertively utters its negation, whatever disagreement exists between them cannot be understood as a disagreement in their beliefs. For he claims that, if we were to focus only on our beliefs, then we could not discern any practical difference between relativist invariantism and non-relativist contextualism.²⁹ And yet MacFarlane holds that relativist invariantism does a better job than non-relativist contextualism in accounting for disagreement. But if only one of these two views can adequately account for disagreement, and if the difference between these views does not arise at the level of belief, then it follows that we must look beyond belief if we are to understand the relevant kind of disagreement.

According to MacFarlane, to find the necessary 'resources for understanding what disagreement looks like for the relativist,' we must consider the 'normative significance of assertions.'³⁰ That is, we must consider the *commitments* we undertake in making assertions. On MacFarlane's view, relativist disagreement can be understood as a conflict among such

²⁸ See MacFarlane [2007] and [unpublished], as well as MacFarlane [manuscript], especially section 5.5.

²⁹ MacFarlane [manuscript], section 5.5.

³⁰ MacFarlane [2007], section 5.2. See also MacFarlane [2005].

commitments.³¹ He describes these commitments as follows. ‘In making an assertion, one licenses others to rely on its accuracy in their actions and reasoning, and one commits oneself to vindicating its accuracy in the face of appropriate challenges.’³² What grounds the possibility of disagreement between speakers occupying different contexts is that, in making an assertion, one undertakes commitments that extend not just to one’s current context, but to *any future context* in which one’s assertion may be appropriately challenged. One’s assertion may be appropriately challenged, according to MacFarlane, whenever the challenger has grounds for thinking that this assertion is false *relative to the context she occupies in issuing the challenge*. And ‘a successful response to such a challenge consists in a demonstration that the assertion was, in fact, accurate (*relative to the context one occupies in giving the response*).’³³

MacFarlane’s account of assertoric commitment appears to provide a satisfying explanation of why it is that, even on a relativist view, Thales and Cavendish disagree when Thales says ‘water might be an element’ and Cavendish says ‘it is not the case that water might be an element.’ For, at least on a sufficiently broad conception of what constitutes a possible context, there will be a possible context, *C*, in which Thales and Cavendish are both appropriately challenged to defend their respective assertions. And so MacFarlane’s account implies that Thales and Cavendish each have the following commitment: if ever *C* should arise, to vindicate the truth of their assertion relative to *C*. But presumably, if ever *C* should arise, it would be impossible for Thales and Cavendish both to succeed in vindicating the truth of their assertions relative to *C*: if either one were to succeed in vindicating the truth of his assertion relative to *C*, then, *ipso facto*, the other would fail. Thus, according to MacFarlane’s theory, in making their respective assertions, Thales and Cavendish take on conflicting commitments. And this fact may explain the disagreement between the two speakers.

³¹ One might get the impression, from MacFarlane [unpublished], that the relevant kind of disagreement is to be understood purely as a logical relation, a relation of incompatibility that he calls ‘preclusion of joint accuracy.’ However, this incompatibility relation cannot account for disagreement unless it is understood in connection with whatever distinguishes assertion from other speech acts. For if Thales *asks* whether water might be an element and Cavendish *asks* whether it is not the case that water might be an element, they don’t thereby disagree. Similarly, if Thales *grants for the sake of argument* that water might be an element, and Cavendish *grants for the sake of argument* the opposite, they don’t thereby disagree. Whatever it is that makes *asserting* things that are ‘precluded from joint accuracy’ out to be a kind of disagreement, therefore, must turn on what differentiates assertion from these other speech acts (compare Schroeder [2008]).

³² MacFarlane [2007, 28].

³³ MacFarlane [2007, 29].

3.3 Why This Account is Incompatible with the Reversibility Thesis

While MacFarlane's theory of assertoric commitment may support the *disagreement* thesis, this theory, when combined with invariantism, is incompatible with the *reversibility* thesis. Recall the case of Rita, who sincerely utters sentence m ('Axeworthy might be, and might not be, the murderer') in the morning, while knowing that, in the evening, she will sincerely utter $\sim m$. Let a_m be the assertion Rita makes in the morning. According to MacFarlane's theory of assertoric commitment, in making this assertion, Rita takes on the following commitment: on the condition that a_m is appropriately challenged in the evening, to vindicate its truth relative to the context Rita occupies in the evening. Now presumably, if making an assertion involves undertaking certain commitments, then making an assertion *sincerely* involves undertaking these commitments *sincerely*, and hence intending to fulfill them. Consequently, MacFarlane's account implies that when Rita sincerely makes assertion a_m in the morning, she must have the following conditional intention: on the condition that a_m is appropriately challenged in the evening, to vindicate the truth of a_m relative to the context she occupies in the evening.

Note that, on the relativist invariantist view, a_m could indeed be appropriately challenged in the evening, because, on this view, a_m has the same truth value as an assertive utterance of m made in the evening, and the latter could clearly be appropriately challenged. Hence, on the invariantist view, Rita could not vindicate the truth of a_m , relative to a given context, without thereby vindicating the claim that an assertive utterance of m made *in the evening* would likewise be true, relative to that same context. For this reason, in the evening, Rita will be both unwilling and unable to vindicate the truth of a_m relative to the context she occupies in the evening. Rita will be *unwilling* to do so because, in the evening, Rita will sincerely assertively utter $\sim m$, and will thus be committed to vindicating this latter assertion, an assertion whose truth is, according to invariantism, incompatible with the truth of a_m . And Rita will be *unable* to vindicate the truth of a_m , relative to the context she occupies in the evening, because it will then be common knowledge whether Axeworthy is the murderer. Hence, on the relativist invariantist view, a_m will be clearly false relative to the context Rita occupies in the evening.

Moreover, Rita should recognize all this in the morning. And if, in the morning, Rita recognizes that, in the evening, she will be both unwilling and unable to vindicate the truth of a_m relative to the context she then occupies, it follows that she could not rationally *intend* to

vindicate the truth of a_m in the face of any appropriate challenge in the evening. But on MacFarlane's theory, as we have seen, Rita must have precisely this intention if she is to sincerely make assertion a_m in the morning. And so MacFarlane's theory of assertoric commitment, together with relativist invariantism, has the unacceptable implication that Rita could not rationally and sincerely make this assertion in the morning.

In response to this difficulty, MacFarlane might move to a weaker conception of the commitments we undertake in making assertions. He might maintain that, in making an assertion, what one undertakes is not a categorical commitment to vindicate its truth in the face of any appropriate challenge, but only a disjunctive commitment to *either vindicate its truth or retract it* in the face of such a challenge.³⁴ Such a move, however, would undermine the pragmatic account of disagreement without successfully accommodating reversibility. It would undermine the pragmatic account of disagreement because, on this view, there is no real conflict between the commitments Thales undertakes in saying 'water might be an element' and the commitments Cavendish undertakes in saying 'it is not the case that water might be an element.' For in any context in which both their assertions are appropriately challenged, they could both fulfill their assertoric commitments by simply retracting. Moreover, this solution does not seem to accommodate reversibility. For on this view, when Rita makes the assertion a_m in the morning, she commits herself to *either vindicating or retracting a_m* in the face of any appropriate challenge in the evening. And so, if her assertion is sincere, she must intend to either vindicate or retract in the face of such a challenge in the evening. But, as we have seen, she knows that in the evening she will be unable to vindicate the truth a_m . And so if she intends the disjunction, she must intend to retract. But if, in asserting p_m in the morning, Rita *recognizes* that this assertion will be indefensible in the face of being challenged in the evening, and if she *already intends to retract* her assertion in the face of any such challenge, then her present assertion would hardly seem to qualify as sincere.

Thus, MacFarlane's theory of assertoric commitment, both in its original form and in the weakened form, fails to reconcile reversibility and disagreement. For the very cross-contextual commitments that allow the theory to account for disagreement make it irrational, on this theory, to sincerely assertively utter a contested sentence when one knows one's assertion will be false relative to future contexts of assessment. Hence, the very features that

³⁴ MacFarlane [2005] includes a commitment to retract among the commitments we undertake in making an assertion. But this is presented as a separate commitment, not as a component of a disjunctive commitment. See also MacFarlane [2007].

allow the pragmatic theory to account for disagreement render it incompatible with reversibility.

4 Rethinking Disagreement

To recap: in part 1 we argued that invariantists are committed to accepting the disagreement thesis, and that *everyone* should accept the reversibility thesis. In part 2 we argued that the invariantist can accept the reversibility thesis only by adopting relativism. And in part 3, we argued that such an adoption of relativism jeopardizes the disagreement thesis, and that to retain this thesis, the relativist needs a pragmatic account of disagreement that is incompatible with the reversibility thesis. Thus, the invariantist cannot accept either of these theses without rejecting the other.

In order to illuminate this dialectic, we will now present what we take to be the most plausible diagnosis of the dilemma facing the invariantist. It is not surprising, we will argue, that the invariantist cannot reconcile these two theses, for these theses stand in deep conflict. The basic problem, we will argue, is that if we accept both these theses, we must conclude that, under standard conditions, one can rationally expect to disagree with one's future self. In the next two sections, we argue that we should reject this conclusion, and with it the claim that we disagree with our future selves in reversibility cases. We then argue, in section 4.3, that this lack of disagreement in reversibility cases undermines the main arguments for invariantism. In the final section, we consider ways in which the invariantist might respond to this problem, and we argue that none can succeed.

4.1 Diagnosis: The Two Theses are Incompatible

Here is an explanation of the difficulties facing the invariantist: she cannot reconcile the disagreement thesis with the reversibility thesis because they are incompatible. And this incompatibility is a problem for the invariantist, for while her opponents can very happily reject the disagreement thesis, she cannot.

The general problem for reconciling the two theses is that disagreement is a kind of *conflict* between individuals. Thus, in order to endorse the disagreement thesis, one must hold that when sincere speakers in arbitrary contexts assert the negations of one another's sentences, there is a conflict between them. But *among* the pairs of contexts in which pairs of

speakers sincerely assert contradictory sentences are pairs that include the same speaker at different times – in situations of reversibility. This means that if the disagreement thesis is correct, then the conflict it predicts must also exist when a single speaker asserts a sentence and its negation at different times. The challenge is to explain how this conflict could be sharp enough to be worth calling ‘disagreement,’ without being sharp enough to make it irrational to stand in this sort of conflict with one’s future self under standard conditions.

The challenge is heightened by the fact that, in ordinary cases of factual disagreement, the kind of conflict that gives rise to the disagreement is the kind that one cannot expect to be in with one’s future self without violating reflection. The invariantist’s hopes ride on whether, in the case of relatively true propositions, it *could* be rational, under standard conditions, to disagree with what one expects to believe and sincerely assert in the future.

And there is reason to doubt that this could be rational. First, under standard conditions, as we have defined them, we expect our future self to be fully rational and to have all the information we presently possess. And there is general reason to doubt the possibility of rationally disagreeing with someone whom one regards as fully rational and as possessing all one’s information.³⁵

Moreover, there is special reason to doubt the possibility of rationally disagreeing with such an individual when she happens to be one’s future self. For if there is genuine, and not merely verbal, disagreement between one’s present and future self, then one should expect that, under appropriate circumstances, this disagreement would manifest itself in a practical conflict. For example, let p be some proposition we now disbelieve but expect to believe in the future, and let C be some possible context where there is the option of betting on p . In such a case, we should expect our future self to conditionally intend to bet on p should she find herself in C , whereas we should now want to prevent our future self from carrying out such an intention.³⁶ Thus, if one genuinely disagrees with what one expects to believe in the future, then one should now prefer to prevent one’s future self from carrying out intentions that it would be rational for

³⁵ Compare Feldman [2006], Elga [2007], and White [2007].

³⁶ The relativist might deny this, and insist that in disbelieving p , we are only committed to wanting that we not bet on p in our *present* context, not to wanting that we not bet on p in *future* contexts. But this move would undermine the claim that there is genuine disagreement between one’s present and future self in the cases under consideration. It would also undermine the claim that it is p (and not the proposition that p is true in our present context) that is the genuine object of belief.

her to have. And, under standard conditions, such a practical conflict with one's future self would appear to be a mark of irrationality.³⁷

And so there is reason to hold that, under standard conditions, *disagreeing with the beliefs* we expect to have in the future is irrational. It does not directly follow that we cannot rationally expect to *disagree* with our future selves. For, as we saw in section 3.2, such disagreement might be understood in pragmatic terms, as a conflict in commitments.³⁸ However – to generalize the conclusion of section 3.3 – invoking such a pragmatic account of disagreement won't help. For the very same considerations that count against the rationality of having beliefs that conflict with one's expected future beliefs also count against the rationality of sincerely undertaking commitments that conflict with the commitments one expects to sincerely undertake in the future. Assuming that sincerely undertaking a commitment involves intending to fulfill this commitment, it follows that such an expected conflict in commitments would involve an expected conflict in intentions: that is, it would involve having intentions that conflict with the intentions one expects to have in the future. And so if this kind of practical conflict with one's future self is irrational under standard conditions, then so is such pragmatic disagreement with one's future self.

Thus, regardless of whether we understand the conflict that constitutes disagreement as a conflict in beliefs or as a conflict in commitments, there is reason to deny that we can rationally expect to disagree with our future self under standard conditions. And since the conjunction of the disagreement thesis and the reversibility thesis entails the opposite, there is reason to regard these theses as incompatible.

4.2 Prognosis: We *Don't* Disagree with our Future Self in Reversibility Cases

If we are right that we cannot rationally expect to disagree with our future self under standard conditions, then it follows that in reversibility cases, we do not disagree with our future self. Fortunately, there is independent reason to accept this conclusion.

To determine whether, in reversibility cases, we disagree with our future self, let us first consider a case where a reversible sentence and its negation are uttered by two *different*

³⁷ This claim is defended in Ross [manuscript].

³⁸ One might object that pragmatic disagreement needn't involve any conflict among beliefs or commitments: it might instead involve only a conflict in pragmatic implicatures or presuppositions. However, if the speakers do not believe, and are not committed to defending, the conflicting implicatures or presuppositions of their assertions, then even if there is a kind of pragmatic conflict between their *speech acts*, there wouldn't seem to be any disagreement between the *speakers*.

speakers. Consider, e.g., a case in which Thales, who is fully rational, utters w ('water might be, and might not be, an element'), and yet he knows that in the distant future, a fully rational scientist named 'Cavendish' will perform an experiment that proves whether water is an element, and that Cavendish will go on to assertively utter $\sim w$. One might plausibly maintain that when Cavendish utters $\sim w$, he disagrees with Thales. For one might imagine Cavendish rationally saying something like the following: 'Thales is wrong about whether water might and might not be an element. For water *can't* be an element, since it consists of hydrogen and oxygen.' Such apparent disagreement, as we saw in section 1.2, underlies one of the central arguments for favoring invariantism over contextualism. But there is an important asymmetry here. For even if *Cavendish* could rationally maintain that *Thales* is wrong about whether water might and might not be an element, it does not seem that *Thales* could rationally maintain that *Cavendish* is wrong about this issue. Indeed, it doesn't seem that Thales could rationally maintain that when Cavendish utters $\sim w$, he is wrong about *anything*. But one cannot disagree with someone without thinking he is wrong about something. Hence, since Thales could not rationally believe that, in saying $\sim w$, Cavendish is wrong about anything, we must conclude that Thales (whom we are assuming to be fully rational) does not disagree with Cavendish.³⁹

Note, however, that in reversibility cases involving epistemic modals, the speaker is initially in a position like that of Thales, not Cavendish. For recall that in reversibility cases, as we have defined them, the speaker does not expect to lose any information between now and the future time when she will utter the negation of what she now utters. Hence, the change in evidence that explains the reversal in what she utters must consist in her *gaining* relevant evidence. For example, what explains why Rita can rationally and sincerely utter m ('Axeworthy might be, and might not be, the murderer'), under standard conditions, while believing that she will later sincerely utter $\sim m$, is that she expects to learn, in the interim, whether Axeworthy is the murderer. And so the same reasoning applies: just as it would be irrational for Thales to regard the better-informed Cavendish as wrong, so it would be

³⁹ Several authors have pointed to similar asymmetries between less- and better-informed speakers (See Wright [2007], Dietz [2008], von Fintel and Gillies [2008], and Björnson and Finlay [manuscript]). However, each of these arguments has involved non-reversible sentences. And such sentences do not conclusively demonstrate that disagreement between less- and better-informed speakers is asymmetric. Consider a non-reversible sentence such as e ('water might be an element'). If Thales knew that Cavendish will sincerely utter $\sim e$ in a context in which Cavendish *knows* that water is not an element, then Thales would thereby know that water is not an element, and so he would not be in a position to assert e . Thales would only be in such a position if he suspected that Cavendish may assert $\sim e$ *without* knowing that water is not an element. And in such a case, it is not obvious that Thales does not disagree with Cavendish.

irrational for Rita to regard her better-informed future self as wrong. Hence, she could not rationally disagree with her future self.

The same arguably goes for reversibility cases involving predicates of personal taste. Suppose Dr. X, who is fully rational, expects that he will perform an experiment that will result in all of humanity acquiring the ability to see the ultraviolet spectrum. He might rationally and sincerely utter d ('dalias are prettier than daisies') while believing that, after the experiment, he will sincerely utter $\sim d$. But he could not rationally believe that, after he, and everyone else, gains greater powers of vision, he will be wrong about which kind flower is prettier. Hence, since Dr. X is fully rational, he must not disagree with his future self.

And the same logic applies in each of the cases of reversibility we discussed in section 1.3. In the lottery case, Winona could not rationally believe that she will later be wrong about the probability that any given ticket is the winning ticket; in Three Envelopes, Mabos could not rationally believe that he will later be wrong about which envelope Chester ought to choose; and in Holmes's Reversal, Holmes could not rationally believe that he will later be wrong about whether any given suspect is the murderer if he had an air rifle. Hence, in each case, the speaker could not rationally disagree with his or her future self. And so if we acknowledge that the speaker could be fully rational in each of these cases, then we must regard each case as a counterexample to the disagreement thesis.

4.3 How this Affects the Debate between Invariantism and Contextualism

Suppose we are right that, in the cases we have been considering, the less-informed speaker does not disagree with the better-informed speaker, so that in reversibility cases we do not disagree with our future self. This conclusion would have important implications for the debate between invariantists and contextualists. First, it would seriously weaken the disagreement-based argument against contextualism. Here's an instance of such an argument:

- (1) Contextualism implies that when Thales assertively utters w ('water might be, and might not be, an element') and Cavendish assertively utters $\sim w$, both assertions could be true.
- (2) But when two speakers make sincere assertions that could both be true, neither speaker thereby disagrees with the other.

- (3) Therefore, contextualism implies that when Thales and Cavendish make these sincere assertions, neither speaker thereby disagrees with the other.
- (4) But when Thales and Cavendish make these sincere assertions, Cavendish thereby disagrees with Thales.
- (5) Therefore contextualism is false.

The crucial premises of this argument are (2) and (4). Now (2) is plausible on the assumption that disagreement between two speakers consists in an incompatibility between their beliefs or assertions, or the impossibility of their beliefs or assertions being jointly true – call this the *incompatibility model of disagreement*. On this model, disagreement must be a symmetric relation: for if x 's beliefs or assertions are incompatible with y 's beliefs or assertions, then the converse must also obtain. But if the argument from the preceding section is sound, then Thales does not disagree with Cavendish. Hence, we must either deny that Cavendish disagrees with Thales, or else we must deny that disagreement is symmetric. If we do the former, we must reject premise (4). And if we do the latter, then we must reject the incompatibility model of disagreement that provides the intuitive support for premise (2). Either way, the argument against contextualism loses much or all of its force.

Similar problems arise for the positive disagreement-based argument in favor of invariantism – namely, that invariantism can explain disagreement in cases where contextualism cannot. For recall that the invariantist explains why Cavendish disagrees with Thales by appealing to the claim that their assertions cannot both be true, and perhaps also by appealing to the claim that their assertions involve undertaking incompatible commitments. But such explanations predict that disagreement should be a symmetric relation: if x 's assertions can't be jointly true with y 's, or involve commitments incompatible with y 's commitments, then the converse also obtains. But we have just seen that we should either reject symmetric accounts of disagreement, or else reject the claim that Cavendish disagrees with Thales. It follows that we should either reject the invariantist explanation of disagreement (since it implies symmetry), or else reject the kind of disagreement claim that invariantism is meant to explain. Either way, we must reject the positive disagreement-based argument for invariantism.

And similar problems arise for the invariantist explanations of retraction and truth and falsity ascription. Recall from section 1.2 that invariantism is motivated, in part, by

observations of the following kind. If Rita were to sincerely utter m , and then subsequently learn whether Axeworthy is the murderer, and sincerely utter $\sim m$, then it would be appropriate for her to retract her first assertion and say that it is false. The invariantist explanation for Rita's retraction and falsity ascription is that Rita's earlier assertion is incompatible with what she later sincerely asserts, and hence with what she later believes. But now consider a case in which, in the morning, Rita is the only one who knows whether Axeworthy is the murderer. But in the evening, she forgets whether Axeworthy is the murderer, and she loses all the relevant evidence, although she remembers that, in the morning, she knew whether he is the murderer. And suppose that, in the morning, Rita sincerely utters $\sim m$, while in the evening, she sincerely utters m . On the invariantist view, the assertion Rita makes in the morning is incompatible with what she sincerely asserts, and hence believes, in the evening. And so the invariantist explanations of retraction and of truth and falsity ascription commit the invariantist to the counterintuitive claim that, in the evening, it would be appropriate for Rita to retract, and regard as false, the assertion she made when he knew whether Axeworthy is the murderer.

And so the invariantist is in a quandary. For her explanations of disagreement and related phenomena are based on an incompatibility model of these phenomena, a model that she has very strong reason to reject. First, as we saw in section 1.2, this model commits her to the disagreement thesis, a thesis that she cannot reconcile with the reversibility thesis. Second, this model commits her to regarding these phenomena as symmetric. But it now appears that either the phenomena of disagreement, falsity ascription and retraction are asymmetric, or else they are absent in the very cases where their presence is supposed to support invariantism.

Perhaps the solution is not to reject invariantism, but rather to combine invariantism with an alternative account of disagreement and related phenomena – an account that does not entail the disagreement thesis, that allows for asymmetries, and that does not imply that we disagree with our future self in reversibility cases. We will now conclude by considering this possibility.

4.4 Can the Invariantist Rethink Disagreement?

We have seen that the invariantist must accept some form of relativism in order to make sense of reversibility. And we have seen that the most straightforward relativist view has the counterintuitive implication that, in cases involving reversible sentences, less-informed

speakers should regard as mistaken speakers whom they acknowledge to be better-informed than them. Perhaps the solution is to adopt a more sophisticated form of relativism. Consider epistemic modal claims. A more sophisticated relativist theory might state that their truth values, relative to a given context of assessment, depend on whatever evidence is *relevant* in that context of assessment. And such a theory might state that the evidence that is *relevant*, in a given context of assessment, includes not only the evidence that is *available* in that context, but also the evidence that is available in the context of utterance of the assertion being evaluated. And so on this view, in the morning, Rita should regard her later utterance of $\sim m$ as true, since it is true relative to the body of evidence that is relevant in the context of assessment she now occupies, a body of evidence that includes both her present evidence and the evidence she will have in the evening.

MacFarlane has proposed precisely this sophisticated relativist view as a response to the apparent asymmetries in truth ascriptions concerning epistemic modal claims.⁴⁰ This view can also be extended to account for corresponding asymmetries in disagreement and retraction. For if we disagree with, or should retract, an assertion just in case we think it's false, then asymmetries in truth ascription will give rise to asymmetries in disagreement and retraction. Hence, on this sophisticated relativist view, we can allow that while Cavendish disagrees with Thales (since Cavendish knows that Thales' assertion is false relative to the inclusive body of evidence that is relevant to Cavendish's assessment), Thales does not disagree with Cavendish (since Thales knows that Cavendish's assertion is *true* relative to the inclusive body of evidence that is relevant to *Thales'* assessment). And, for similar reasons, we can allow that while Rita should retract her earlier assertive utterance of m in the case where she has subsequently *learned* whether Axeworthy is the murderer, she should not retract her earlier assertive utterance of $\sim m$ in the case where she has subsequently *forgotten* whether Axeworthy is the murderer. This view also implies that, in reversibility cases, the speaker does not disagree with her future self. For this view implies that, in such cases, the speaker's later assertion is true relative to the inclusive body of evidence that is relevant in her initial context of assessment.

Unfortunately, this sophisticated relativist view has unacceptable consequences. For in order to explain why Cavendish should regard Thales' assertive utterance of w as false, we need to assume that the evidence that is relevant in evaluating a given assertion includes the evidence available to the *assessor*. And in order to explain why Thales should not regard

⁴⁰ MacFarlane [forthcoming, 38-41].

Cavendish's assertive utterance of $\sim w$ as false, we need to assume that the evidence that is relevant in evaluating a given assertion includes the evidence available to the *speaker*. But together these two assumptions have highly counterintuitive implications. Suppose Pozzo is in Potsdam, and Lucky is in London. Suppose Pozzo and Lucky both recognize that Pozzo knows how hot it is in Potsdam, but not how hot it is in London, and that Lucky knows how hot it is in London, but not how hot it is in Potsdam. And suppose Pozzo says: 'It might be hotter in London than in Potsdam and it might not be.' The view under consideration has the absurd consequence that Lucky should regard Pozzo's assertion as false, since it is false relative to any body of evidence that includes both Lucky's evidence and Pozzo's evidence.

Moreover, even if some sophisticated relativist strategy could solve the above problem, no such strategy can accommodate reversibility without sacrificing invariantism. The general strategy is to explain why the less-informed speaker does not disagree with the better-informed speaker, in the cases we have been considering, by adopting a kind of relativism on which the better-informed speaker's assertion is true relative to the less-informed speaker's context of assessment. Hence, the proponent of this strategy must claim that, in Old News, Rita's later assertive utterance of $\sim m$ is true, relative to the context of assessment she occupies in the morning. However, since Rita is in a position to assertively utter m , rather than $\sim m$, in the morning, an assertive utterance of $\sim m$, *made in the context of utterance Rita occupies in the morning*, cannot be true as assessed from the context she occupies in the morning. And so anyone who adopts this general strategy must claim that, relative to the context of assessment Rita occupies in the morning, the truth value of an assertive utterance of $\sim m$ depends on its context of utterance. The proponent of this strategy must therefore reject invariantism, and adopt some form of contextualism (perhaps relativist contextualism).⁴¹

Furthermore, the proponent of this strategy will be driven to a contextualism of a very radical kind. For she wants to maintain that when a less-informed and a better-informed speaker utter the negations of one another's contested utterances, the better-informed speaker should hold that the less-informed speaker is wrong, while the less-informed speaker should not hold that the better-informed speaker is wrong about anything. But now consider the case in which, in the morning, Rita utters m , and then adds 'and that's the truth.' Then, in the evening, after learning whether Axeworthy is the murderer, she utters $\sim m$, and then retracts

⁴¹ *Pace* MacFarlane, who claims that this sophisticated relativist view avoids being overly 'concessive to contextualism' [forthcoming, section 8].

her earlier assertive utterance of *m*, calling it false. In this case, the proponent of the sophisticated relativist strategy must say that it is perfectly appropriate for Rita to say, in the morning, that her present assertive utterance of *m* is *true*, without believing that she will be wrong when, in the evening, she says that this assertion is *false*. But this could only be appropriate if the truth values of truth ascriptions depend on their context of utterance. Hence, the proponent of this strategy is committed to contextualism about truth itself.

It appears, therefore, that if we take at face value the claims about disagreement, retraction and truth ascription that motivate invariantism, and we also acknowledge that, in reversibility cases, the less-informed speaker does not disagree with the better-informed speaker, then we cannot accept invariantism. We must either accept a thoroughgoing contextualism, or else dispense altogether with propositional accounts of the contested expressions.

References

- Bach, Kent [forthcoming]. 'Perspectives on Possibilities: Contextualism, Relativism, or What?' Forthcoming in Egan and Weatherson, eds., *Epistemic Modals*. Oxford: Oxford University Press.
- Björnson and Finlay [manuscript]. 'Metaethical Contextualism Defended.' Unpublished paper.
- Cappelen, Herman, and Ernest Lepore [2004]. *Insensitive Semantics: A Defense of Semantic Minimalism and Speech Act Pluralism*. Oxford: Blackwell.
- Chrisman, Matthew [2007]. 'From Epistemic Contextualism to Epistemic Expressivism.' *Philosophical Studies* 135(2): 225-254.
- Christensen, David [1991]. 'Clever Bookies and Coherent Beliefs.' *Philosophical Review* 100(2): 229-247.
- Dietz, Richard [2008]. 'Epistemic Modals and Correct Disagreement.' In García-Carpintero and Kölbel, eds., *Relative Truth*. Oxford: Oxford University Press.
- Egan, Andy [2006]. 'Secondary Qualities and Self-Location.' *Philosophy and Phenomenological Research* 72(1): 97-119.
- _____. [2007]. 'Epistemic Modals, Relativism, and Assertion.' *Philosophical Studies* 133(1): 1-22.

- _____. [2010]. ‘De Gustibus Non Disputandum Est (At Least, Not Always).’ Forthcoming in Feldman and Warfield, eds., *Disagreement*. Oxford: Oxford University Press.
- Egan, Andy, John Hawthorne, and Brian Weatherson [2005]. ‘Epistemic Modals in Context.’ in G. Preyer and P. Peter, eds., *Contextualism in Philosophy*. Oxford: Oxford University Press.
- Elga, Adam [2007]. ‘Reflection and Disagreement.’ *Noûs* 41(3): 478-502.
- Feldman, Richard [2006]. ‘Epistemological Puzzles about Disagreement.’ In Stephen Heatherington, ed., *Epistemology Futures*. Oxford: Oxford University Press, 194-214.
- Gibbard, Allen [1981]. ‘Two Recent Theories of Conditionals.’ In William Harper, Robert Stalnaker, and G. Pearce, eds., *Ifs: Conditionals, Beliefs, Decisions, Chance, Time*. Dordrecht: Reidel.
- _____. [1990]. *Wise Choices, Apt Feelings*. Cambridge: Harvard University Press.
- Gillies, Anthony [2004]. ‘Epistemic Conditionals and Conditional Epistemics.’ *Noûs* 38(4): 585-616.
- Kölbel, Max [2002]. *Truth Without Objectivity*. New York: Routledge.
- Kolodny, Niko, and John MacFarlane [unpublished a]. ‘Ought: Between Subjective and Objective.’ Unpublished manuscript.
- Kolodny, Niko, and John MacFarlane [unpublished b]. ‘Ifs and Oughts.’ Unpublished Manuscript.
- Laserson, Peter [2005]. ‘Context dependence, disagreement, and predicates of personal taste.’ *Linguistics and Philosophy* 28: 643-686.
- Lewis, David [1979]. ‘Attitudes *De Dicto* and *De Se*.’ *Philosophical Review* 88(4): 513-543.
- MacFarlane, John [2003]. ‘Future Contingents and Relative Truth.’ *Philosophical Quarterly* 53: 321-336.
- _____. [2005]. ‘Making Sense of Relativist Truth.’ *Proceedings of the Aristotelian Society* 105: 321-339.
- _____. [2007]. ‘Relativism and Disagreement.’ *Philosophical Studies* 132: 17-31.
- _____. [2008]. ‘Truth in the Garden of Forking Paths.’ In M. Kölbel and M. Garcia-Carpintero, eds., *Relative Truth*. Oxford: Oxford University Press, 81-102.
- _____. [2009]. ‘Nonindexical Contextualism.’ *Synthese* 166: 231-250.
- _____. [forthcoming]. ‘Epistemic Modals are Assessment-Sensitive.’ Forthcoming in Weatherson and Egan, eds., *Epistemic Modality*, Oxford: Oxford University Press.

- _____. [unpublished]. 'Varieties of Disagreement.' Unpublished paper, draft of 6/18/09, available at <http://johnmacfarlane.net/varieties.pdf>.
- _____. [manuscript]. *Assessment Sensitivity: Relative Truth and Its Applications*. Unpublished partial book manuscript, draft of 12/13/09, available at <http://johnmacfarlane.net/retruth-chs1-7.pdf>.
- Parfit, Derek [forthcoming]. *On What Matters*. Forthcoming from Oxford University Press.
- Price, Huw [1983]. 'Does 'Probably' Modify Sense?' *Australasian Journal of Philosophy* 61: 396-408.
- Regan, Donald [1980]. *Utilitarianism and Cooperation*. Oxford: Oxford University Press.
- Richard, Mark [2004]. 'Contextualism and Relativism.' *Philosophical Studies* 119: 215-242.
- Ross, Jacob [2006]. *Acceptance and Practical Reason*, Doctoral Dissertation, Rutgers University.
- _____. [manuscript]. 'Reasons, Rationality, and Commitment.' Unpublished paper.
- Schroeder, Mark [2008]. *Being For: Evaluating the Semantic Program of Expressivism*. Oxford: Oxford University Press.
- Stanley, Jason [2005]. *Knowledge and Practical Interests*. Oxford: Oxford University Press.
- Stephenson, Tamina [2007 a]. 'Judge Dependence, Epistemic Modals, and Predicates of Personal Taste.' *Linguistics and Philosophy* 30(4):
- _____. [2007 b]. 'Indicative Conditionals Have Relative Truth Conditions.' *Proceedings of the Annual Meeting of the Chicago Linguistics Society* 43(1): 231-242.
- Stevenson, C. L. 'The Emotive Meaning of Ethical Terms' (1937). Reprinted in C. L. Stevenson. *Facts and Values*. Westport, CT: Greenwood Press, 1963. 10-31.
- van Fraassen, Bas [1984]. 'Belief and the Will'. *Journal of Philosophy* 81: 235-256.
- Veltman, Frank [1996]. 'Defaults in Update Semantics.' *Journal of Philosophical Logic* 25(3): 221-261.
- von Fintel, Kai, and Anthony Gillies. 'CIA Leaks.' *Philosophical Review* 117(1): 77-98.
- Weatherson, Brian [2009]. 'Conditionals and Indexical Relativism.' *Synthese* 166: 333-357.
- White, Roger [2005]. 'Epistemic Permissiveness.' *Philosophical Perspectives* 19: *Epistemology*, 445-459.
- Wright, Crispin [2007]. New Age Relativism and Epistemic Possibility: The Question of Evidence. *Philosophical Issues* 17 :262-283.
- Yalcin, Seth [forthcoming]. 'Nonfactualism About Epistemic Modality.'