Reversibility or Disagreement

Jacob Ross
University of Southern California
jacobmro@usc.edu

Mark Schroeder
University of Southern California
maschroe@usc.edu

The phenomenon of disagreement has been recently brought into focus by the debate between contextualists and relativist invariantists about epistemic expressions such as 'might', 'probably', indicative conditionals, and the deontic 'ought'. Against the orthodox contextualist view, it has been argued that an invariantist account can better explain apparent disagreements across contexts by appeal to incompatibility of the propositions expressed in those contexts. This paper introduces an important and underappreciated phenomenon associated with epistemic expressions – a phenomenon that we call *reversibility*. We argue that the invariantist account of disagreement is incompatible with reversibility, and we go on to show that reversible sentences cast doubt on the putative data about disagreement, even without assuming invariantism. Our argument therefore undermines much of the motivation for invariantism, and provides a new source for constraints on the proper explanation of purported data about disagreement.

Contextualists, about a given type of expression, claim that the semantic contribution of an expression of this type can depend on its context of utterance. Invariantists, on the other hand, deny this. Recently, arguments have been given favoring invariantism for a number of expressions for which contextualism had previously been the dominant view, including particularly evidence-dependent expressions such as epistemic modals ('might', 'must'),[1] deontic modals ('ought'),[2] probability operators ('probably', 'as likely as not'),[3] and indicative conditionals,[4] among others.[5] For brevity, we will call such expressions *epistemic expressions*.

Invariantists have prominently argued that their view is preferable to contextualism because only their view accounts for disagreement and closely related phenomena. For invariantism seems to allow, seemingly correctly, that Thales and Cavendish disagree when Thales says 'water might be an element' and Cavendish, upon discovering the composition of water, says 'it is not the case that water might be an

---

[1] Egan, *et al* (2005), MacFarlane (2011), Stephenson (2007a).
[2] Kolodny and MacFarlane (unpublished).
[3] Egan (2007), MacFarlane (2011).
[4] Weatherson (2009), Stephenson (2007b), Kolodny and MacFarlane (2010).
[5] In addition to the evidence-dependent or 'epistemic' expressions that we will discuss in this paper, a similar dialectic has arisen for predicates of personal taste such as 'tasty' and 'funny' (see Lasersohn (2005) and Stephenson (2007a)). 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)).

element.' Contextualism, on the other hand, seems to predict that they do not thereby disagree, since the proposition Thales affirms differs from, and is consistent with, the proposition Cavendish denies. Such disagreement claims have provided one of the central motivations for invariantism about epistemic expressions.

A much less appreciated phenomenon about epistemic expressions is that sometimes sentences involving such expressions are *reversible*. Consider the case of *Old News*.

> *Old News*: Ankita is the anchorwoman for the *Morning News Hour*. As of this morning, it is unknown whether Axeworthy is the murderer, and so Ankita assertively utters *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. Ankita also knows that the anchorman for the *Evening News Hour* is very sloppy, and she expects that this evening he will assertively utter *m*. And she correctly predicts that when she hears this, she will exclaim: 'Nonsense! It is not the case that Axeworthy might be, and might not be, the murderer.'

Here the sentence *m* has the peculiar feature that Ankita can rationally and sincerely assertively utter it while believing that later, when she is better informed, she will rationally and sincerely assertively utter its negation. We call this feature *reversibility*.

Our paper will be concerned with the relationship between reversibility and disagreement. In particular, we will argue that reversibility is incompatible with the conception of disagreement that is needed if disagreement is to provide support for invariantism. Thus, the phenomenon of reversibility undermines much of the motivation for the invariantist view.

Our main argument, laid out in parts 2 and 3, is that invariantists face a dilemma: they can account for reversibility, or they can account for disagreement, but they can't do both. An introductory part 1 establishes concepts and terminology and sets the stage, and a concluding part 4 concerns the question of how we must rethink disagreement in light of the phenomenon of reversibility.

1   Stage-Setting

In this section we'll begin by defining our terms for purposes of this paper, before going on to clarify the theses about disagreement and reversibility that will figure in the remainder of the paper.

1.1   A Taxonomy of Views

In order to focus more clearly on the kinds of sentences for which the invariantist is able to predict disagreement, let us define an *epistemic sentence* as a sentence containing one of the epistemic expressions

under dispute ('might', 'probably', etc.) but no other components which make a contextually varying contribution to the content or truth conditions of what is asserted. And let us define an *epistemic proposition* as a proposition expressed by an assertive utterance of an epistemic sentence.[6]

As one example of an epistemic sentence, consider 'water might be an element'. There are several views one might take about such a sentence. On a *propositional* view, the semantic contribution of an epistemic sentence is a proposition, understood as something which determines truth-conditions for the sentence.[7] In contrast, on a *non-propositional* view, such as expressivism, this is not the case. In this paper, we will be concerned only with propositional views; at each point in the argument we will simply be assuming for convenience that non-propositional views are off of the table.[8] In assuming this for convenience, we do not mean to be assuming that there are any conclusive arguments against non-propositional theories, but merely to be limiting the theoretical options for purpose of this discussion.

Propositional views can be divided into what we may call 'contextualist' and 'invariantist' views. Contextualists affirm, while invariantists deny, that when an epistemic expression is uttered, its semantic contribution (i.e., its contribution to the content or to the truth conditions of what is semantically expressed) can depend on the context of utterance. Since we have defined an *epistemic sentence* as one containing an epistemic expression but no other context-sensitive components, it follows that on the contextualist view, the context in which an epistemic sentence is uttered can affect either the proposition it semantically expresses or the truth value of this proposition. By contrast, on the invariantist view, no such variation is possible in the case of epistemic sentences.

Another exhaustive distinction among propositional views is the distinction between what we may call 'relativist' views and 'absolutist' views. Absolutists maintain that there is a perspective-independent answer as to whether a given epistemic proposition is true, or as to whether someone spoke truly in asserting such a proposition. By contrast, the relativist maintains that the truth value of an epistemic proposition, or of an assertive utterance thereof, depends on the context *from which it is evaluated*. Note that we are defining our terms differently from those who treat 'contextualism,' 'relativism,' and 'invariantism' as

---

[6] The point of these definitions is to allow us to isolate the contextual contribution of epistemic expressions, and thereby substantially simplify the discussion. It may turn out to be impossible to completely isolate the contextual contribution of epistemic expressions in this way in natural language, in which case our definition relies on an idealization. But we believe that all of the main points in this paper could be made – much more laboriously – without this idealization, by carefully controlling for other possible differences in context. So we will ignore this possibly very important qualification in what follows.

[7] For the importance of this qualification, see Schroeder (forthcoming).

[8] Expressivist views are most familiar in the moral case, but may also be defended for epistemic modals (Price (1983), Yalcin (2007), (2011)), 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). For convenience, we also classify Kent Bach's radical invariantism (Bach 2011) as a non-propositional theory, though Bach's view has much in common with contextualism.

mutually exclusive and jointly exhaustive kinds of propositional views. Our taxonomy has the advantage of making it clear that there are really two orthogonal questions: one concerning variation with the context of utterance, and another concerning variation with the context of assessment. Because these questions are orthogonal, they mark out four, not three, possible kinds of view, which our terminology allows.

One final term that we should clarify is "sincerely." As we will be using this term, *x sincerely assertively utters s* just in case:

(i)   *x* assertively utters *s* with the intention to assert the proposition that is semantically expressed by this utterance; and
(ii) *x* believes that her assertion is true.

Our use of "sincerely" can thus be thought of as short for "sincerely and with the intention to assert what is semantically expressed." We adopt this usage simply in order to save words.

I.2   The Disagreement Thesis

Disagreement has long been central to the debate between contextualists and invariantists.[9] And it has seemed to pose a serious problem for contextualism. 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 *consistent* 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 *inconsistent* with what is known in *his* context of utterance. And since water's being an element could be consistent with what is known in one context and inconsistent with what is known in another, the simple contextualist view implies that their assertions are *compatible* in the sense that they could both be true, and so 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 this particular case. But *any* contextualist theory will predict that there are *some* cases in which one speaker assertively utters one epistemic sentence in one context, and another speaker assertively utters its negation in another context, and the two assertions are compatible. Thus, any version of contextualism will be difficult to reconcile with the thesis that speakers like Thales and Cavendish disagree in general. By contrast, invariantism would seem to be well suited to explaining why such speakers would disagree, since this view implies that their assertions *cannot* both be true.

---

[9] This dialectic isn't new; compare the criticisms of speaker subjectivism in ethics in Stevenson (1937).

Note that this explanation of Thales and Cavendish's putative disagreement explains it in terms of the incompatibility of their beliefs. Call this *disagreement by incompatibility*. Any such explanation of disagreement is going to be committed to the following:

> *disagreement thesis*: For any epistemic sentence, *s*, and competent speakers, *x* and *y*, if *x* sincerely assertively utters *s* in one context, and *y* sincerely assertively utters ∼*s* in another context, then *x* and *y* thereby disagree.[10]

Disagreement has not, of course, been the only way in which invariantism about epistemic expressions has been motivated. It has also been argued that invariantism does better than contextualism in relation to some other phenomena that are related to but not identical with disagreement, including data about retraction and about truth and falsity ascriptions.

Suppose that before the crucial experiment is performed, Cavendish says 'water might be an element,' and then he subsequently discovers that water is a compound. Invariantists have argued that after learning this, it makes sense for him to 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 Cavendish is now in a position to make an assertion that is incompatible with his former assertion. Contextualism, however, seems to predict the opposite, since on this view, given the context in which Cavendish's original assertion was made, he should still regard what he said as true, and so it would seem that he should stand by it.[11]

Since the problem of disagreement has been central to the debate between contextualists and invariantists, it will be the main focus of our dilemma for invariantists in parts 2 and 3. However, in part 4 we will argue that some of the problems we raise for the invariantist account of disagreement extend to the invariantist accounts of retraction and of truth and falsity ascriptions.

1.3   The Reversibility Thesis

We began the paper with the case of *Old News*, which described a case of *rational reversibility*. In this section we will further clarify what we mean by saying that some epistemic sentences are reversible, and show that

---

[10] 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). (Note that Egan (2010) is focusing on predicates of personal taste, rather than on epistemic expressions.; Egan defends analogous views about epistemic modals elsewhere.)

[11] 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 (2010). And for a discussion of some relevant empirical data, see Yalcin and Knobe (2010).

each epistemic expression (epistemic modals, probability operators, indicative conditionals, and the deontic 'ought') figures in some reversible epistemic sentences. In particular, we will defend the following thesis:

> *reversibility thesis*:   For each kind of epistemic expression, there are epistemic sentences involving that kind of expression that a fully rational speaker can sincerely assertively utter, even under ideal conditions, while correctly believing that later she will sincerely assertively utter their negations.

By *ideal conditions*, we mean conditions in which the speaker is fully confident that her relevant beliefs will evolve in a way that is rationally required in virtue of her acquiring additional evidence. It will be important in section 2.1 that these are conditions under which the principle of reflection governs rational belief. Indeed, the reversibility thesis is itself closely related to the principle of *reflection*, and that close relationship is exactly what will be important to our argument.

It can be easily shown by example that the reversibility thesis is true, for reversible sentences can be constructed involving each kind of epistemic expression. We've already illustrated a reversibility case for 'might' with *Old News*, and the duality of 'might' and 'must' allows the same case to work for 'must'. In *Old News*, it is safe to assume that ideal conditions obtain. That is, we may assume that Ankita is fully confident that her relevant beliefs will change in a manner that is rationally required by her acquiring additional evidence, namely, the results of the DNA test. Even on this assumption, Ankita could rationally and sincerely assertively utter m in the morning while correctly believing that, in the evening, she will assertively utter $\sim m$. And so this case instantiates the reversibility thesis.

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 choose 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 in the second envelope, then when they rewatch the film he will say: 'It's 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's not the case that Chester ought to choose the first envelope. For he ought to choose the third.'[12]

---

[12] This example is from Ross (2006). It is structurally analogous to the mine shaft problem introduced in Regan (1980), and discussed in Parfit (2011) and in Kolodny and MacFarlane (2010).

Once again, even on the assumption that ideal conditions obtain, Three Envelopes is a case in which Mabos can rationally and sincerely assertively utter an epistemic sentence while correctly believing he will later sincerely assertively utter its negation.

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

> *Holmes's Reversal:* Moriarty, Milverton, and Moran are the three possible suspects for 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 *a*.) However, Holmes also knows that he will soon learn the identity of the innocent air-rifle owner, and consequently that he will know, of each of the remaining two suspects, that if he had the air rifle, then he is the murderer. And so Holmes correctly predicts that, at this later time, he will sincerely and assertively utter $\sim a$.

Again, even on the assumption that ideal conditions obtain in this case, Holmes can rationally anticipate a reversal in what he sincerely assertively utters. And so indicative conditionals, like epistemic and deontic modals, instantiate the reversibility thesis. So, too, do probability operators. Under ideal conditions, 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 correctly believing that, after the winning ticket has been revealed, she will rationally and sincerely utter $\sim l$.

Thus, in each of the contested domains, there are sentences that are rationally reversible even under ideal conditions, as the reversibility thesis states.

There are a few clarificatory remarks to make about the reversibility thesis before proceeding. First, it's important to note that the reversibility thesis concerns *epistemic* sentences, which are defined as sentences that include epistemic expressions and *no other context-sensitive components*. If we were to omit this qualification, and simply state that, for every epistemic expression, there is a sentence in which it figures that a rational speaker can sincerely assertively utter while correctly believing that she will later sincerely assertively utter its negation, then the resulting thesis would be trivial, since it would be true of any expression: we can construct such a sentence involving any expression, simply by introducing other context-sensitive expressions.

Second, it's important to clarify the significance of the clause, 'even under ideal conditions'. Once again, without this clause the reversibility thesis would cease to be interesting, for what it would say about epistemic sentences would then be true of just about any sentence whatsoever. For there are almost always

non-ideal conditions under which one could rationally and sincerely assertively utter a sentence while correctly believing that later one will sincerely utter its negation. All that needs to be true for this to happen is for you to believe that you will either irrationally change your mind, or rationally change your mind as a result of losing information. But while just about any sentence can be reversible under non-ideal conditions, what is distinctive about certain epistemic sentences is that they are reversible even under ideal conditions.

It is also important to note that *not all* sentences involving epistemic expressions are reversible. Thus, while sentences of the form *it might be that p and it might be that not p* are typically reversible, simpler epistemic sentences of the form *it might be that p* are never reversible, at least for ordinary (non-epistemic) values of *p*. Consider, for example, sentence *m** ('Axeworthy might be the murderer'). Under ideal conditions, Ankita could not rationally and sincerely assertively utter *m** while believing that she will later sincerely assertively utter ~*m**. For in this case she would have to believe that she will later have conclusive evidence that Axeworthy is not the murderer, and if she really believed that there will be such conclusive evidence, then she would not be in a position to rationally and sincerely assertively utter *m** now. Similarly, while we have seen that some complex sentences involving probability operators and indicative conditionals are reversible, simpler sentences of the form *probably p* or *if p then q* are never reversible, for ordinary values of *p* and *q*. There are, however, some syntactically simple sentences that are reversible. 'Chester ought to take the first envelope' is one example, as we have seen. 'It is uncertain whether Axeworthy is the murderer' is another example. In general, reversibility arises when there is some knowledge-base *B* and some question *Q* such that

(i) Having precisely knowledge-base *B* is sufficient for being in a position to correctly assertively utter *s*;
(ii) Having any knowledge-base derived from enriching *B* with *any* answer to *Q* is sufficient for being in a position to correctly assertively utter ~*s*.

Finally, the phenomenon of reversibility would appear to be related to the more familiar phenomenon that has been discussed in the literature under the name "faultless disagreement." One might wonder, therefore, why we are focusing on reversibility, and not on this more familiar phenomenon. In particular, one might wonder (and, indeed, an anonymous referee has wondered) why we base our arguments on the reversibility thesis, and not instead on the following thesis:

Apparent Faultless Disagreement (AFD): There is a contested sentence *s* that a fully rational speaker *x* could sincerely assertively utter at some point of time *t*, under standard conditions, while there is a

point of time t' and a speaker y such that x believes that y sincerely utters the negation of s at t' and x cannot rationally maintain that y is wrong in making this utterance.

We have two main reasons for favoring the reversibility thesis over AFD as a basis for our argument. First, while we expect that something like AFD is true, we don't think it provides as uncontroversial a starting point for argumentation as the reversibility thesis, since the invariantist might deny AFD. After all, on the invariantist view, we should regard those who assert the negations of our epistemic sentences as asserting something *false*, and so the invariantist might maintain that we should regard such speakers as *wrong* in making these assertions. Moreover, there are some relativists who would reject this formulation of AFD since they would maintain that wrongness must itself be understood relative to a context of assessment. Second, there is an important respect in which AFD is too weak for our purposes. While AFD entails that cases can arise in which a rational speaker would believe that *someone* assertively utters the negation of what she herself assertively utters, AFD does not entail that cases can arise in which a rational speaker would believe that *she herself will later* assertively utter the negation of what she now assertively utters. Thus, AFD fails to entail that there can be *intrapersonal , diachronic cases* in which the phenomenon in question arises. And yet some of our key arguments turn on the existence of such intrapersonal, diachronic case, and so we need a principle that entails their existence.[13]

2   The Problem of Reversibility

Having argued that the reversibility thesis is true, we will now discuss its significance for invariantism. In section 2.1, we will argue that because of the principle of *reflection*, non-relativist versions of invariantism are incompatible with reversibility. Then in section 2.2, we will argue that, by contrast, relativist versions of invariantism are compatible with relativism. In section 2.3, we will consider an alternative way in which the invariantist might attempt to account for reversibility, and we will argue that it fails, and that only relativism can succeed. This conclusion will set the stage for part 3, where we argue that the invariantist who accepts the reversibility thesis, and thus accepts relativism, cannot endorse the disagreement thesis.

---

[13] In response to this second objection, the anonymous referee has replied that "AFD may be strengthened to a claim to the effect that for any given pair of speakers and utterance times, it is (at least metaphysically) possible for them to provide an instance of AFD." However, if AFD were strengthened in this way, then it would be clearly false, since it would fail in the case where $x = y$ and $t = t'$. In general, any version of AFD that fails to imply the existence of intrapersonal, diachronic cases will be, in that respect, too weak for our purposes, and any principle that implies the existence of additional cases will be, in that respect, stronger than we need.

2.1 Why Absolutist Invariantism is Incompatible with Reversibility

Let us consider a case of reversibility. Consider, in particular, the case in which, under ideal conditions, Ankita sincerely assertively utters m ('Axeworthy might be and might not be the murder') in the morning while believing that in the evening she will sincerely assertively utter $\sim m$. And let us assume, for the sake of *reductio*, that Ankita, who is fully rational, believes the conjunction of invariantism and absolutism. And so she denies that the truth value of an assertive utterance of an epistemic sentence depends on the context in which it is uttered, and she likewise denies that the truth value of such an utterance depends on the context from which it is evaluated.

Now if someone sincerely assertively utters some sentence, she must believe that her utterance is true, at least as evaluated from the context she occupies in making the assertion. Hence, since, in the morning, Ankita sincerely assertively utters m ('Axeworthy might be, and might not be, the murderer'), it follows that, in the morning, Ankita believes the following:

> T_1: An assertive utterance of m made in the context of utterance Ankita occupies in the morning is true relative to the context of assessment Ankita occupies in the morning.

Since Ankita believes T_1 together with invariantism, and since she is fully rational, she must believe the following consequence of these beliefs.

> T_2: An assertive utterance of m made in the context of utterance Ankita occupies in the *evening* is true relative to the context of assessment Ankita occupies in the morning.

Since Ankita believes T_2 together with absolutism, and since she is fully rational, she must believe the following consequence.

> T_3: An assertive utterance of m made in the context of utterance Ankita occupies in the evening is true relative to the context of assessment Ankita occupies in the *evening*.

Recall, however, that in the morning, Ankita believes that in the evening she will sincerely assertively utter $\sim m$. Hence, in the morning, Ankita must believe that, in the evening, she will believe the following.

> T_4: An assertive utterance of $\sim m$ made in the context of utterance Ankita occupies in the evening is true relative to the context of assessment Ankita occupies in the evening.

But T4 entails the negation of T_3. Thus, in the morning, since Ankita believes that in the evening she will believe T_4, and since she expects her beliefs to evolve in a rational manner, she must believe that in the

evening she will *disbelieve* T_3. Hence, in the morning, Ankita must believe T_3 while believing that in the evening she will disbelieve that same proposition.

It follows that Ankita violates the principle of reflection. Following van Fraassen, where p represents a proposition, Cr_1 represents an agent's credence function at an earlier time, and Cr_2 represents this same agent's credence function at a later time, let us say that this agent *satisfies reflection* with respect to p just in case,

Reflection: $Cr_1(p \mid Cr_2(p = x)) = x$ for all x.

That is, she satisfies reflection with respect to p just in case her credence in p, conditional on the supposition that her future credence in p will be x, is itself x. If Ankita believes T_3 in the morning while believing that in the evening she will disbelieve it, then she will violate reflection with respect to T_3. For if a fully rational agent believes a given proposition, then presumably she must have high credence in it. And if she believes that she will later disbelieve it, then presumably she must have high credence that she will later have low credence in it. But if one has high credence in a proposition, and at the same time one has high credence that one will later have low credence in it, and if one's credences are probabilistically coherent, then for some values of x,

$$Cr_1(p \mid Cr_2(p = x)) > x$$

where Cr_1 represents one's credences at the earlier time at Cr_1 represents one's credences at the later time. Hence, one will violate reflection with respect to this proposition. And so it follows that Ankita, whom we have stipulated to be fully rational, violates reflection with respect to T_3.

But this conclusion is not plausible, and so it constitutes a *reductio* of our initial assumptions. For we have assumed that ideal conditions obtain, and hence that Ankita is fully confident that her relevant beliefs will evolve in a way that is rationally required in virtue of her acquiring additional evidence. And ideal conditions are precisely the sorts of conditions under which we should expect a fully rational agent to satisfy *reflection* with respect to a proposition such as T_3—while there are well-known cases in which a fully rational agent can violate reflection with respect to such a proposition, these all involve non-ideal conditions, e.g., cases involving anticipated irrationality or information loss.[14] Since ideal conditions obtain

---

[14] For discussions of such cases, see Christensen [1991], Bovens [1995], Elga [2000], Arntzenius [2003] and Briggs [2009].

in the case we are considering, we should deny that Ankita can rationally violate reflection with respect to T3. And so we should reject the conclusion we have just derived.

But this derivation proceeded from the assumption that the absolutist invariantist theory can be accepted by a fully rational agent who is in a reversibility situation. It doesn't follow, strictly speaking, that the absolutist invariantist theory is *false*. One might hold that this theory is true, but that no one can rationally accept it in reversibility situations. But this view, according to which the correct semantic theory of our language cannot be rationally accepted by its users in ordinary contexts of use, is hardly plausible. And so we may conclude that, in light of reversibility, absolutist invariantism is not a tenable position.

2.2   Why Relativist Invariantism is Compatible with Reversibility

Relativist invariantism avoids the problem for absolutist invariantism discussed in the previous section. For suppose we begin with the premise that Ankita accepts relativist invariantism rather than absolutist invariantism. In this case, while Ankita will still believe T_1 (since she sincerely assertively utters m in the morning) and hence she also believes T_2 (since she is an invariantist), we can block the inference that Ankita believes T_3 (since she is a relativist). Thus, in virtue of her accepting relativism, she can maintain that while an assertive utterance of m made in the evening would be true relative to the context of assessment she occupies in the morning, such an assertive utterance would not be true relative to the context of assessment she will occupy in the evening. Hence, there needn't be any disagreement between Ankita's present beliefs and her anticipated future beliefs concerning T_3. And so Ankita needn't violate reflection with respect to T_3.

Admittedly, even on the relativist view, Ankita will violate reflection in the present case. She won't violate reflection with respect to T_3, but she will violate reflection with respect to another proposition, namely the proposition that is semantically expressed by assertive utterances of m—call this proposition p_m. Since Ankita sincerely assertively utters m in the morning, while anticipating that in the evening she will sincerely assertively utter the negation of m, she must believe p_m in the morning while believing that in the evening she will disbelieve p_m. And, for reasons already given, this means that she will violate reflection with respect to p_m.

It can be argued, however, that if relativism is true, then while the supposition that Ankita can rationally violate reflection with respect to T_3 is highly problematic, the supposition that she can rationally violate reflection with respect to p_m is unproblematic. For if relativism is true, then p_m is not an ordinary

proposition: it is not the sort of proposition whose truth conditions can simply be identified with a set of possible worlds. Rather, it is a proposition of a very special kind, one whose truth conditions must be relativised to contexts of assessment. By contrast, T_3 is an ordinary proposition whose truth conditions can be identified with a set of possible worlds. While T_3 is *about* context-relative truth, its own truth conditions are not *themselves* context-relative. Now it seems that, under ideal conditions, fully rational agents satisfy reflection with respect to ordinary propositions. And so we should expect such agents to satisfy reflection with respect to T_3. But the situation is very different with respect to a non-standard proposition such as p_m. If we expand our conception of propositions to include non-standard propositions, then we should expect rational violations of reflection.

Consider, for example, the view that, in addition to ordinary, "uncentered" propositions, there are also *centered propositions*. On this view, while an *uncentered* proposition determines a set of *worlds* at which it is true, a *centered* proposition determines a set of *centered worlds* at which it is true, where a centered world is a triple of the form $<w, t, i>$, consisting in a possible world w, a privileged time t, and a privileged individual i.[15] To believe a centered proposition corresponding to a given set of centered worlds is to believe that one of the worlds in that set represents where, when, and who one is. Thus, to believe the centered proposition (call it *April 1ˢᵗ*) associated with the set of centered worlds in which the privileged time is April 1ˢᵗ, 2012 is to believe oneself to be located at that time. Now clearly, on April 1ˢᵗ, 2012, under ideal conditions, someone could rationally locate herself at April 1ˢᵗ, 2012, while believing that the next day she will locate herself at a different time. Hence it follows, on the view under consideration, that someone could rationally believe *April 1ˢᵗ* while believing that the next day she will believe ~*April 1ˢᵗ*. Hence, if we expand our conception of propositions to include centered propositions, then we must maintain that our beliefs in these propositions are not governed by the principle of reflection, even under ideal conditions.

Now there is a close connection between centered propositions and assessment-sensitive propositions, that is, propositions whose truth values depend on their context of assessment. The connection is this: if we recognize both kinds of proposition, then we should hold that, for every assessment-sensitive proposition p, there is a centered proposition p^* such that it is rational for one to believe p iff it is rational for one to believe p^*. In particular, for any assessment-sensitive proposition p, it will be rational to believe p iff it's rational to believe the centered proposition believing which constitutes believing oneself to be located in a context of evaluation relative to which p is true. But if it's rational to believe a given assessment-sensitive proposition just in case it's rational to believe the corresponding

---

[15] See Lewis [1979].

centered proposition, and if, as we have seen, one can rationally violate reflection with respect to such centered propositions, then it follows that one can rationally violate reflection with respect to assessment-sensitive propositions. Thus, if we expand our conception of propositions to include assessment-sensitive propositions, then we should expect such propositions to lie outside the scope of the principle of reflection.

Let us sum up the argument from the last two sections. In order to avoid the implausible implication that fully rational agents can violate reflection, under ideal conditions, with respect to ordinary propositions such as T_3, the invariantist needs to adopt relativism. The resulting view will still imply that reversibility cases involve rational violations of reflection, but this implication will be unproblematic, since it will concern a proposition to which we should not expect the principle of reflection to apply. And so relativist invariantism, unlike absolutist invariantism, can be reconciled with reversibility.

2.3   But…is Tense Doing Work?

In the last two sections we have argued that, in order for the invariantist to account for reversibility, she must accept relativism. But one very distracting feature of each of our examples of reversibility is that they unfold over time. And that makes it very natural to wonder whether tense, rather than the presence of epistemic expressions, plays some important role in the examples. One might wonder, for example, whether the epistemic component of epistemic sentences makes an absolute and invariant contribution to the truth of sentences, but each of these sentences has *another* component – tense – which makes a contextually variant contribution, and whether this is what is doing the work in our examples.

Unfortunately, however, there are at least three straightforward reasons why we do not believe that this idea provides any way around our argument. The first is simple: many of the examples used to motivate invariantism, such as the example involving the apparent disagreement between Thales and Cavendish, concerns utterances made at different times. If examples like that motivate the view that their respective contexts do not affect the truth value of their assertions, then they equally motivate the view that their respective *times* do not affect the truth value of their assertions, and so if these sentences are tensed, then by the lights of the invariantist that still cannot affect their truth value.

This should not be a surprise, because in general (and this is our second reason why this idea cannot save absolute invariantism), tensed sentences do not shift in truth value over time unless there is some other expression in the sentence to which time is relevant. For example, 'Twice two is four' is tensed, but does not vary in truth across times, because time is not relevant for elementary arithmetic. In contrast,

'Caroline has been born' shifts in truth value across times, but only because the tense shifts the time parameter relevant for 'born'. The same goes for epistemic expressions. If time is not relevant for their interpretation, then adding a tense, which sets or shifts the time parameter for the sentence or clause, will not change anything.

Third and most importantly, while standard cases of reversibility involve assertions occurring at two different times, such separation in time is not essential to the underlying phenomenon. Consider, for example, a case in which Ankita assertively utters m at 9 am, while believing that at noon she will find out whether Axeworthy is the murderer, and that she will then travel through a wormhole and arrive on the other side of town just before 9 am, whereupon she will assertively utter $\sim m$ at 9 am. In this case, what explains how Ankita can rationally anticipate such a reversal in what she assertively utters can't be the tensed character of m. For the time at which Ankita expects to assertively utter $\sim m$ is the same as the time at which she assertively utters m. Thus, difference in time is an incidental feature of the most realistic examples of reversibility, and not central to the phenomenon itself.

3   The Problem of Disagreement

So far, we have argued that because of the truth of the reversibility thesis, invariantism can be plausible only if it is combined with relativism. In what follows, we will use this result in order to press our dilemma for the invariantist. We begin, in section 3.1, by arguing for a *prima facie* incompatibility between relativism and the disagreement thesis. We then argue that, in order for the relativist to plausibly resist our argument and to defend the compatibility between her position and the disagreement thesis, she must adopt a *pragmatic* account of disagreement. We consider two such accounts: Egan's account in section 3.2 and MacFarlane's account in section 3.3. We argue that Egan's account does not succeed in reconciling relativism with disagreement. And we argue that, while MacFarlane's account does succeed in reconciling relativism with disagreement, it does so only at the cost of sacrificing reversibility. Hence, adopting relativism does not enable the relativist to endorse both the disagreement thesis and the reversibility thesis after all; it simply moves the invariantist from one horn of our dilemma to the other.

3.1   The *Prima Facie* Incompatibility between Relativism and Disagreement

Recall that we defined the disagreement thesis as follows:

disagreement thesis: For any epistemic sentence, *s*, and competent speakers, *x* and *y*, if *x* sincerely assertively utters *s* in one context and *y* sincerely assertively utters ∼*s* in another context, then *x* and *y* thereby disagree.

As we saw in section 1.2, this thesis plays an important role in motivating invariantism, and it is meant to state, as a general principle, an important phenomenon that contextualists can't account for. Thus, the word 'disagree' that figures in this thesis must be understood in the ordinary, pretheoretic sense in which it is supposedly a datum that disagreement exists in cases where contextualism implies that it does not. Our question is therefore whether, if relativism is true, sincere utterers of contradictory epistemic sentences always stand in a relation that we would ordinarily describe as disagreement.

We will now argue that, at least on the face of it, relativism appears to be incompatible with the disagreement thesis understood in this manner. To see why this is so, it will be useful to first consider the version of relativism defended by Egan. As we saw in section 2.2, if the relativist view of epistemic propositions is correct, then these propositions are closely related to *centered* propositions. According to Egan's version of relativism, this connection is very close indeed, for on his view, epistemic propositions *are* centered propositions. According to Egan,

> The semantic value of 'Bond might be in Zurich' (is) the set of all $<w, t, i>$ 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.[16]

Similarly, on Egan's view, when a speaker assertively utters sentence *e* ('water might be an element'), she asserts a centered proposition that locates her in a situation in which water's being an element is consistent with all the available evidence. Thus, if Thales sincerely assertively utters *e* and Cavendish sincerely assertively utters ∼*e*, then Thales expresses a self-locating belief whereby he locates himself in one kind of situation, and Cavendish expresses a self-locating belief whereby he locates himself in a different kind of situation.

But we know independently that two people who locate themselves in different situations 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, even implicitly. Nor do they disagree when Thales locates himself in a situation in which fresh fish are readily available, and Cavendish locates himself in a context in which fresh fish are hard to come by. More generally, the following appears to be true:

---

[16] Egan (2007, 8).

16

> ***I'm OK , You're OK***: If a first person is in one kind of context, and a second person is in another kind of context, then they could have self-locating beliefs whereby the first person correctly locates herself in a context of the first kind, and the second person correctly locates himself in a context of the second kind, without disagreeing explicitly or implicitly.[17]

It follows that Thales and Cavendish will not disagree by virtue of the fact that Thales locates himself in a context where the available evidence is *consistent* with water's being an element, and Cavendish locates himself in a context where the available evidence is *inconsistent* with water's being an element. But on Egan's view, it is precisely these self-locating beliefs that they express when Thales sincerely assertively utters *e* and Cavendish sincerely assertively utters ~*e*. Hence, if we assume that speakers disagree in making their respective assertions only if they disagree in having the beliefs these assertions express (later we will discharge this assumption), then it follows, on Egan's view, that there needn't be any disagreement between Thales and Cavendish when they sincerely assertively utter *e* and ~*e*, respectively, contrary to the disagreement thesis.

Thus, although Egan is committed to the disagreement thesis,[18] his theory appears, at least *prima facie*, to be incompatible with it. For, on Egan's theory, to believe an epistemic modal proposition *just is* to have a certain self-locating belief, and the self-locating beliefs expressed when two people sincerely assert the negations of one another's epistemic sentences patently do not involve disagreement. We will now argue that this very same problem generalizes to other relativist views. For, while some relativists may not accept Egan's identification of beliefs in epistemic propositions with self-locating beliefs, on any plausible relativist view, the former beliefs will be closely connected with the latter. More precisely, on any plausible relativist view, the following will be true:

> ***Rational Commitment***: The self-locating belief whereby one locates oneself in a context relative to which *p* is true rationally commits one to believing *p* (in the sense that a perfectly rational agent cannot have the former belief without the latter) and, similarly, the self-locating belief whereby one locates oneself in a context relative to which ~*p* is true rationally commits one to believing ~*p*.[19]

---

[17] The distinction between explicit and implicit disagreement can be illustrated by the following pair of cases. If I believe that Socrates is immortal and you believe that Socrates is mortal, then we *disagree explicitly*, since our beliefs are in direct opposition. Suppose, however, that while I believe that Socrates is immortal, you have never considered the question of Socrates' mortality. Suppose, however, that you believe that Socrates is a man, and you also believe that all men are mortal. In this case, we *disagree implicitly*. For, if you were to reason correctly on the basis of your beliefs, and thereby arrive at the conclusions to which they commit you, then you would believe that Socrates is mortal, and so we would come to have directly opposing beliefs.

[18] See, e.g., Egan (2010), where he says that speakers like Thales and Cavendish are engaged in a 'genuine conflict.'

[19] Our relativist targets appreciate the importance of this principle; see See Egan *et al* (2005), MacFarlane (2005, 333) and especially MacFarlane (manuscript), section 5 of chapter 5 on belief, most explicitly on page 139.

On the basis of this Rational Commitment principle, we will now argue that, on a relativist view where *p* can be true relative to one context and ~*p* can be true relative to another, if one person in one context believes *p* and another person in another context believes ~*p*, then they needn't thereby disagree, contrary to the disagreement thesis. The argument proceeds as follows. Assume that relativism is true. Hence,

A1. From the assumption of relativism: for some epistemic proposition *p*, it is possible for one person to be in a context relative to which *p* is true, and for another person to be in another context relative to which ~*p* is true.

A2. From A1 and *I'm OK, You're OK*: for some epistemic proposition *p*, it's possible for two people in different contexts to have self-locating beliefs whereby the first correctly locates herself in a context relative to which *p* is true, and the second correctly locates himself in a context relative to which ~*p* is true, without disagreeing explicitly or implicitly.

A3. From A2 and *Rational Commitment*: for some epistemic proposition, *p*, it's possible for two people in different contexts to be in mental states that rationally commit them to believing *p*, and to believing ~*p*, respectively, without disagreeing explicitly or implicitly.

A4. For any two propositions, *p* and *q*, if it's impossible for two people in different contexts to believe *p* and *q*, respectively, without disagreeing, then it's likewise impossible for two people in different contexts to be in mental states that rationally commit them to believing *p*, and to believing *q*, respectively, without disagreeing at least implicitly.

A5. From A3 and A4, by *modus tollens*: for some epistemic proposition *p*, it's possible for two people in different contexts to believe *p* and ~*p*, respectively, without disagreeing.

A6. For any epistemic propositions, *p* and *q*, if it's possible for two people to believe *p* and *q*, respectively, without disagreeing, then it's likewise possible for them to sincerely assert *p* and *q*, by way of assertively uttering any pair of sentences expressing *p* and *q*, respectively, without disagreeing.

A7. From the assumption of invariantism and our definitions of 'epistemic proposition': for any epistemic proposition, *p*, there is some epistemic sentence *s*, such that assertive utterances of *s* always express *p*, and assertive utterances of ~*s* always express ~*p*.

A8. From A5, A6 and A7: for some epistemic proposition, *p*, and some sentence, *s*, it is possible for two people in different contexts to assert *p* and ~*p*, respectively, by assertively uttering *s* and ~*s*, respectively, without disagreeing – contrary to the disagreement thesis.

Thus, from the assumption that relativist invariantism is true, along with some other *prima facie* plausible premises, we have inferred that the disagreement thesis is false. In the next two sections, we will examine two ways in which the relativist might attempt to resist this argument's conclusion by rejecting one of its premises. Both of these lines of resistance involve adopting a *pragmatic account of disagreement*. Adopting one of these accounts (the one proposed by Andy Egan) will involve rejecting premise A6, while adopting the other of these accounts (the one proposed by John MacFarlane) will involve rejecting premise A4.

3.2   Egan's Pragmatic Account of Disagreement, and Why it Fails

Andy Egan has an account of assertion and an account of disagreement which together entail the negation of premise A6. These can be stated as follows.

> **Belief-Transfer Model of Assertion:** What one must believe in order to *accept* an assertion is the same as what one must believe in order to sincerely *make* the assertion, namely the proposition that is its content.[20]

> **Non-cotenability Model of Disagreement:** When a first speaker sincerely makes a first assertion and a second speaker sincerely makes a second assertion, they thereby disagree just in case neither party could consistently make his or her assertion while accepting the other's assertion.[21]

Assuming, with Egan, that there are such things as centered propositions, it's easy to see that these two principles jointly conflict with A6. To see how, consider the centered proposition believing which constitutes believing oneself to be located in Greece, that is, the centered proposition corresponding to the set of all centered worlds $<w, t, i>$ such that, in world w, individual i is in Greece at time t. Call this centered proposition *Greece*, and call its negation *~Greece*. As noted earlier, Thales could locate himself in Greece and Cavendish could locate himself outside of Greece without disagreeing. Hence, given our assumption that self-locating beliefs are beliefs in centered propositions, it follows that Thales could believe *Greece* and Canvendish could believe *~Greece* without disagreeing. But it follows from the above two accounts that they could not *assert* these two propositions, respectively, without disagreeing. For according to the Non-cotenability Model of Disagreement, they could assert these propositions without disagreeing only if, in so doing, they could each consistency accept the other's assertion. And according to the Belief-Transfer Model of Assertion, this will be true only if each speaker could simultaneously believe the proposition his assertion expresses and the proposition the other speaker's assertion expresses. But clearly no one could consistently believe *Greece* (thereby locating himself in Greece) while simultaneously believing *~Greece* (thereby locating himself outside of Greece). And so it follows from Egan's accounts of assertion and of disagreement that Thales and Cavendish couldn't assert *Greece* and *~Greece*, respectively, without disagreeing, even though they could believe *Greece* and *~Greece*, respectively, without disagreeing. And so these accounts are jointly inconsistent with A6.

---

[20] See especially Egan (2007). See also Egan (2010). We borrow the name from MacFarlane (unpublished).
[21] Egan (2010).

Egan would of course acknowledge that Thales and Cavendish could sincerely assertively utter "I'm in Greece" and "I'm not in Greece" without disagreeing. What this shows, according to Egan, is that assertive utterances of "I'm in Greece" and of "I'm not in Greece" *don't* express the centered propositions *Greece* and ∼*Greece*, respectively. Rather, they express respectively, the uncentered proposition that Thales is in Greece and the uncentered proposition that Cavendish is not in Greece—since it is the former proposition that Cavendish must believe in order to accept Thales assertion, and it is the latter proposition that Thales must believe in order to accept Cavendish's assertion. Hence, according to Egan, there is a very important difference between indexical sentences and epistemic sentences. Assertive utterances of indexical sentences typically express uncentered propositions, and which such proposition they express depends on the context of utterance. It is for this reason that people in different contexts can assertively utter the negations of one another's indexical sentences without disagreeing.

By contrast, assertive utterances of epistemic sentences express centered propositions, and they express the same centered proposition regardless of the context of utterance. Thus, assertive utterances of 'water might be an element' express the proposition believing which constitutes locating oneself in a context in which all the available evidence is consistent with water's being an element. Since one person can locate herself in such a context, and another person can locate herself outside such a context, without disagreeing, Thales can believe this centered proposition and Cavendish can believe its negation without disagreeing. But since neither party could consistently believe the centered propositions expressed by both of these assertions simultaneously, it follows from Egan's account of disagreement that, if Thales and Cavendish were to assert these propositions, they would thereby disagree.

The problem with this account of disagreement, however, is that it makes disagreement too cheap. To see why, let us suppose that Cavendish has not yet discovered that water is an element, and so he, like Thales, assertively utters "water might be an element." And suppose, further, that there is no other respect in which he disagrees with Thales. Suppose that this lack of disagreement disheartens Jambi, the time travelling genie, who likes conflict. Fortunately, while reading Andy Egan's works, Jambi has thought of a way to create disagreement between Thales and Cavendish. He will do so by expanding their idiolects. Jambi first pays a visit to Thales, who is hanging out with his friend Anaxagoras, and Jambi instructs them as follows:

> In order to assert that proposition believing which constitutes locating oneself in Greece, say "Meka leka hi meka hiney ho." And in order to assert that proposition believing which constitutes locating oneself outside of Greece, say "Meka leka hi meka *chiney* ho." These sentences are importantly different from the sentences "I'm in Greece" and "I'm not in Greece." For if person A

20

assertively utters "meka leka hi meka hiney ho," then another person B should accept A's assertion only if *B* is in Greece. Hence, since in making an assertion you aim for your conversation-partners to accept it, it is felicitous to assertively utter "meka leka hi meka hiney ho" only if the speaker and all her conversation partners are in Greece.[22] Similarly, it's felicitous to assertively utter "meka leka hi meka chiney ho" only if the speaker and all her conversation partners are outside of Greece.

Thales and Anaximander catch on right away, assertively uttering "meka leka hi meka *hiney* ho" to one another, and nodding in agreement. Jambi then travels forward in time to pay a visit to Cavendish and his buddy Lavoisier, to whom he introduces the same expressions, providing the same explanations. They too learn quickly, and assertively utter "meka leka hi meka *chiney* ho" to one another. Being very close the border of Greece, and wanting to try out the other new expression they have learned, they enter Greece and assertively utter "meka leka hi meka *hiney* ho" to one another. They then leave Greece and once again assertively utter "meka leka hi meka *chiney* ho."

Egan's theory implies that when Thales assertively utters "meka leka hi meka *hiney* ho" and Cavendish assertively utters "meka leka hi meka *chiney* ho," they disagree with one another, or stands in "genuine conflict," since neither one could consistently make his own assertion while accepting the other's assertion. Egan's theory also implies that Cavendish comes in and out of disagreement with Thales when he moves back and forth across the Greek border, and hence alternates between the two kinds of utterance. But these implications are hardly plausible. For it's hardly plausible that Thales and Cavendish genuinely disagree simply in virtue of expressing their self-locating beliefs using the vocabulary provided to them by Jambi. Nor is it plausible that they come in and out of disagreement, in predictable ways, simply in virtue of Cavendish's going back and forth across the Greek border and using his new vocabulary to express his changing location. But note that, on Egan's theory, the kind of opposition that exists between Thales and Cavendish when they assertively utter *e* ("water might be an element") and ~*e* respectively, is exactly the same as the opposition that exists between them when they assertively utter "meka leka hi meka *hiney* ho" and "meka leka hi meka *chiney* ho," respectively. And so if we were to accept Egan's they, then we should deny that Thales and Cavendish genuinely disagree when they assertively utter *e* and ~*e* respectively, and so we should reject the disagreement thesis.

---

[22] This injunction corresponds to the "presupposition of relevant similarity" discussed in MacFarlane (2007).

3.3 MacFarlane's Pragmatic Account of Disagreement, and Why it Fails

So far we have considered one way in which the relativist might resist the argument of section 3.1, namely by rejecting premise A6. While we have argued that this response is unsuccessful, there is another response to our argument that may be more promising. This response consists in rejecting premise A4. Just as the negation of A6 can be derived from Egan's account of assertoric disagreement (disagreement involving assertions), so the negation of A4 can be derived from MacFarlane's account of doxastic disagreement (disagreement involving beliefs), which can be stated as follows.

> *Preclusion of Joint Accuracy*: Two people disagree just in case they have a pair of beliefs such that for any context of assessment, *C*, as assessed from *C* these beliefs can't both be true or accurate.[23]

We can derive the negation of A4 if we accept this account of disagreement, and if we also accept a contrast that MacFarlane draws between two kinds of propositions: assessment-sensitive propositions and explicitly relativized propositions. On MacFarlane's view, where *p* is an assessment-sensitive proposition, while the truth value of *p* will vary with the context of assessment, the truth value of the proposition *p is true relative to context of assessment C* will not itself vary with the context of assessment.

Suppose all this is correct. And suppose MacFarlane is right that epistemic propositions are assessment-sensitive. Accordingly, let *p* be some epistemic proposition, and suppose that, at time *t*, *p* is true relative to the context of assessment occupied by some individual *i*, and ~*p* is true relative to the context of assessment occupied by some other individual *j*. And suppose that, at *t*, *i* has the following three beliefs.

I1: The belief in *p*.
I2: The belief that *p* is true relative to the context of assessment occupied by *i* at *t*.
I3: The self-locating belief whereby i locates herself at a context of assessment relative to which *p* is true.

Suppose further that, at *t*, *j* has the following three beliefs.

J1: The belief in ~*p*.
J2: The belief that ~*p* is true relative to the context of assessment occupied by *j* at *t*.
J3: The self-locating belief whereby *j* locates himself at a context of assessment relative to which ~*p* is true.

---

[23] See MacFarlane (unpublished).

Since proposition p is assessment-sensitive, the truth of I1 and of J1 will vary with the context of assessment. Relative to i's context, I1 will be true or acurate and J1 will be false or inaccurate, and relative to j's context, J1 will be true or acurate and I1 will be false or inaccurate. But there will be no context of assessment relative to which both these beliefs are true or accurate. And so it follows from the Preclusion of Joint Accuracy account that i and j can't have beliefs I1 and J1, respectively, without disagreeing. By contrast, since I2 and J2 are beliefs in explicitly relativized propositions, their truth value does not vary with the context of assessment. Each one is absolutely true, or true relative to any context of assessment. And I3 is true or accurate just in case I2 is true or accurate, and similarly J3 is true or accurate just in case J2 is true or accurate. Hence, I3 are J3 are both true relative to any context of assessment. And so it follows from the Preclusion of Joint Accuracy account that i and j can have beliefs I3 and J3, respectively, without disagreeing. But these beliefs, which i and j *can* have without disagreeing, commit them to beliefs I1 and J1, respectively, which they *can't* have without disagreeing. And so it follows that the present case is a counterexample to A4, given MacFarlane's assumptions.

But in order for this argument to have any force in undermining A4, we need some reason to accept MacFarlane's claim that, on the relativist picture, "preclusion of joint accuracy" would amount to genuine disagreement. In particular, we need some explanation as to why there would be any genuine conflict or disagreement between an individual i who believes p and an individual j who believes $\sim p$, when both parties recognize that p is true relative to i's context and $\sim p$ is true relative to j's context. If these parties agree about which propositions are true relative to each context of assessment, then how can there be any real disagreement between them? Fortunately, MacFarlane has an answer to this question. The basic idea is to explain the disagreement involved in having these beliefs derivatively, in terms of the pragmatic conflict that would arise were these speakers to express these beliefs by making assertions.

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.'[24] That is, we must consider the *commitments* we undertake in making assertions.[25] On MacFarlane's view, relativist disagreement can be understood as a conflict among such commitments. 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.'[26] What grounds the possibility of disagreement between speakers occupying different contexts is that in making an assertion,

---

[24] MacFarlane (2007), section 5.2. See also MacFarlane (2005).
[25] Compare MacFarlane (2005).
[26] MacFarlane (2007, 28).

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*).'[27]

MacFarlane's account of assertoric commitment appears to provide a satisfying explanation of how it can be the case, even on a relativist view, that 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, according to relativist invariantism, there is no context of assessment relative to which both Thales's and Cavendish's assertions are true. Hence, if either one succeeded, in *C*, in vindicating the truth of his assertion relative to *C*, then the other would *ipso facto* fail. Thus, MacFarlane's theory implies that in making their respective assertions, Thales and Cavendish undertake conflicting commitments. And this fact may explain how there can exist a genuine, though pragmatic, form of disagreement between the two speakers.

But there's a problem. While MacFarlane's theory of assertoric commitment may support the *disagreement* thesis, this theory, when combined with relativist invariantism, is incompatible with the *reversibility* thesis. Recall the case of Ankita, 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 ~*m*. Let a_m be the assertion Ankita makes in the morning. According to MacFarlane's theory of assertoric commitment, in making this assertion, Ankita 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 Ankita 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 Ankita 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.

---

[27] MacFarlane (2007, 29).

Note, however, that, on the relativist invariantist view, a_m could indeed be appropriately challenged in the evening, since as assessed from the context of someone issuing a challenge in the evening, a_m is false. Moreover, in the evening, Ankita will be both unwilling and unable to vindicate the truth of a_m relative to the context she then occupies. Ankita will be *unwilling* to do so because in the evening, she will sincerely assertively utter $\sim m$, and so, on the relativist invariantist view, she'll be committed to vindicating the truth of an incompatible assertion. And Ankita will be *unable* to vindicate the truth of a_m, relative to the context she occupies in the evening, because it will then be commonly known whether Axeworthy is the murderer. And so, on the relativist invariantist view, a_m will be clearly false relative to the context Ankita occupies in the evening. And Ankita should recognize all these facts in the morning. And so when she assertively utters m in the morning, she should recognize that she will be both unwilling and unable to vindicate the truth of this assertion in the evening.

But one cannot rationally intend to do what one knows one will be both unwilling and unable to do.[28] Hence, in the morning, Ankita 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, Ankita must have precisely this intention if she is to sincerely make assertion a_m in the morning. And so MacFarlane's theory has the unacceptable implication that Ankita could not rationally and sincerely make this assertion in the morning. And the same logic will apply in any case of reversibility: in any such case, MacFarlane's theory will imply that the speaker cannot rationally and sincerely assertively utter the epistemic sentence in question when she is aware that she will later sincerely assertively utter its negation. And so his theory of assertoric commitment, combined with his relativist view of epistemic sentences, is incompatible with the reversibility thesis.[29]

Thus, MacFarlane's theory of assertoric commitment fails to reconcile reversibility and disagreement. For the very claims about cross-contextual commitments that can allow the theory to

---

[28] Note that here, as in section 2.1, our argument turns on an important and deeply plausible assumption about how a rational agent's current attitudes must relate to her expectations about her future situation. In both cases, the tension with these compelling principles comes from diachronic cases of intrapersonal reversibility — as characterized by our formulation of the reversibility thesis. Though there may be something analogous to rational reversibility in interpersonal cases — as indeed we will go on to argue in part 4 — its existence is therefore irrelevant for both of our main arguments about reversibility.

[29] 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 vindicating 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. (Note that MacFarlane (2005) includes a commitment to retract among the commitments we undertake in making assertions. But this is presented as a separate commitment, not as a component of a disjunctive commitment. See also MacFarlane (2007).) Such a disjunctive move, however, would undermine the pragmatic account of disagreement. For on this revised 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.

account for disagreement imply that it would be irrational to sincerely assertively utter an epistemic sentence when one knows one's assertion will be false relative to future contexts of assessment. Hence, the very features that allow the pragmatic theory to account for disagreement render it incompatible with reversibility.[30]

4   Rethinking Disagreement

To recap: in part 1 we explained the important role that the disagreement thesis plays in motivating invariantism, and argued that *everyone* should accept the reversibility thesis. In part 2 we argued that the invariantists 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 after all. Thus, the invariantist cannot accept either of these theses without rejecting the other. This completes our argument for the dilemma facing the invariantist.

In order to shed more light on this dilemma, however, we will now present what we take to be the most plausible diagnosis of its source. It is not surprising, we will argue, that the invariantist cannot reconcile the disagreement thesis and the reversibility thesis, for reversible sentences are general counterexamples to the disagreement thesis. Consequently, the reversibility thesis and the disagreement thesis stand in conflict, quite independently of invariantism. And since, as we have argued, the reversibility thesis is undeniable, this means we must reject the disagreement thesis—again, quite independently of whether we accept invariantism. Thus, what is problematic about invariantism is not that the invariantist cannot reconcile these two theses (since no one can do so) but rather that while her opponents can happily reject the disagreement thesis, the invariantist cannot easily do so without undermining much of the motivation for her view. And so the tension between these theses creates a special problem for the invariantist.

In the next section, we argue that cases of reversibility are indeed counterexamples to the disagreement thesis. Then in section 4.2, we show how the rejection of the disagreement thesis undermines the main arguments for invariantism. And finally in section 4.3 we consider a way in which the invariantist might attempt to salvage these arguments without endorsing the disagreement thesis, and argue that it fails.

---

[30] We should note that MacFarlane's account of assertoric commitments is arguably better-suited to its original application – to the case of future contingents – than to epistemic expressions, since future contingents don't involve rational reversibility. See MacFarlane (2003).

4.1   Why Reversibility Cases are Counterexamples to the Disagreement Thesis

Our argument for the claim that we do not disagree with our future self in reversibility cases turns on an important asymmetry that was first pointed out by Richard Dietz (2008). Dietz considers a case in which, on Monday, Naomi is in a context where it is not yet known whether there are any counterexamples to Goldbach's conjecture, and so she assertively utters sentence g ('there might be a counterexample to Goldbach's conjecture'). But by Tuesday, Naomi has discovered a proof of Goldbach's conjecture, and so she assertively utters $\sim g$. On the standard relativist picture, there is some proposition p_g such that

> B1.   Naomi believes and asserts p_g on Monday, and she believes and asserts $\sim p_g$ on Tuesday.
>
> B2.   When Naomi believes and asserts p_g, she thereby disagrees with anyone who believes and asserts $\sim p_g$.
>
> B3. p_g is true relative to the context Naomi occupies on Monday, and $\sim p_g$ is true relative to the context she occupies on Tuesday.

This view implies that, as assessed from the context Naomi occupies on Tuesday, on Tuesday she not only *disagrees* with her earlier, less-informed self, but further she *correctly disagrees* with her earlier self. For the belief and assertion she makes on Tuesday (according to B1) constitutes disagreement with her earlier self (according to B2) and is correct (according to B3). And this implication may seem fairly plausible.

However, as Dietz points out, the standard relativist view has another implication that is far less plausible. It implies that, as assessed from the context Naomi occupies on Monday, on Monday she correctly disagrees with her future, better-informed self, since she correctly believes and asserts the negation of what she will later believe and assert. But this doesn't seem right. In the kind of case under consideration, when a less-informed speaker and a better-informed speaker assert the negations of one another's epistemic sentences, it doesn't seem that the less-informed speaker could correctly regard the better-informed speaker's assertion as wrong. And so, even as assessed from the context of the less-informed speaker, it doesn't seem that she could correctly disagree with the better-informed speaker. We may state Dietz's claim in general form as follows:

> ***Dietz's Principle***: For any epistemic sentence, s, any competent speakers, x and y, and any times, t_1 and t_2, such that y at t_2 is strictly better informed than x at t_1, if x assertively utters s at t_1, and y assertively utters $\sim s$ at t_2, and if each of these assertions is true as assessed from the context in which it is made, then, even as assessed from the context x occupies at t_1, y's assertion is *not wrong* and so this is not a case in which x at t_1 correctly disagrees with y at t_2.[31][32]

---

[31] Dietz (2008) focuses on epistemic modal sentences, but the point naturally generalizes to other kinds of epistemic sentence.

We regard Dietz's principle, and Dietz's associated argument against the standard form of relativism, as very compelling. And we are not alone in this assessment. MacFarlane himself concedes the force of this argument and acknowledges that it requires a revision to the standard relativist view.[33] Our present concern, however, is with the disagreement thesis, and so we must go one step beyond Dietz's principle. What we must argue for is not the claim that, in some cases where speakers assert the negations of one other's epistemic sentences, the less-informed speaker does not disagree *correctly* with the better-informed speaker, but rather the stronger claim that in some such cases, the less-informed speaker does not disagree *at all* with the better-informed speaker. We will now use Dietz's principle (or, more precisely, the claim that a fully rational agent is in a position to recognize the truth of this principle), in order to argue for this stronger conclusion.

To provide such an argument, we must consider a case involving a reversible sentence. Let us consider, in particular, the *Old News* case, where in the morning Ankita, who is fully rational, assertively utters m ('Axeworthy might be, and might not be, the murderer'), while believing that in the evening she will assertively utter $\sim m$. In this case, the following claims appear to be true in the morning when Ankita assertively utters m:

C1.   Ankita is in a position to recognize the truth of Dietz's principle.

C2.   Ankita is in a position to recognize that, while her present assertive utterance of m is true as assessed from her present context, her assertive utterance of $\sim m$ will be true as assessed from the context she will occupy in the evening when she is better informed.

C3.   Thus, from C1 and C2, Ankita should recognize that her future assertion is not wrong, even as assessed from her present context.

C4.   Thus, since Ankita is fully rational, she will *in fact* recognize that her future assertion is not wrong as assessed from her present context.

C5.   For any competent speakers, x and y, and any sentences s_1 and s_2, and any times t_1 and t_2, if x can rationally and sincerely assertively utter s_1 at t_1 while recognizing that, even as assessed from x's present context, it would not be wrong for y to sincerely assertively utter s_2 at t_2, then the two speakers can make these sincere assertions without x at t_1 disagreeing with y at t_2.

---

[32] What is the sense of 'wrong' at issue here? As MacFarlane (forthcoming) points out, the word 'wrong' can be used in many ways. All that is required for our present argument, however, is that the relevant sense of 'wrong' is one on which judgments of wrongness are connected to disagreement, so as to make premise C5 (below) come out true.

[33] See MacFarlane (2011). Dietz's argument is also endorsed in Wright (2007) and in Björnson and Finlay (2010).

C6. Therefore, Ankita can sincerely assertively utter *m* in the morning, and ~*m* in the evening, without its being the case that, in the morning, Ankita disagrees with her future self – contrary to the disagreement thesis.

Note that the feature of *Old News* on which this argument rests is the following: there is some sentence, *s*, which a fully rational speaker assertively utters while believing that later, when she is better informed, she will assertively utter its negation, and her future utterance will be true, at least as assessed from the context in which it is made. And this feature is present in *every* reversibility case involving epistemic sentences. And so if the argument is sound, then we may generalize, and conclude that in any such reversibility case, where a fully rational speaker sincerely assertively utters one epistemic sentence while correctly believing that she will later sincerely assertively utter its negation, she does not thereby disagree with her future self.

4.2   How the Failure of Disagreement Threatens to Undermine Invariantism

As we saw in section 1.2, the most prominent arguments for favoring invariantism over contextualism are that the invariantist is said to provide a better account of disagreement, and of related phenomena including retraction and truth and falsity ascriptions. Thus if, in the morning, Ankita assertively utters *m* ('Axeworthy might be, and might not be, the murderer'), and then in the evening she assertively utters ~*m*, then invariantism implies that her two assertions are incompatible, in the sense that they cannot both be true (relative to any given context of assessment). Hence, invariantism seems to imply, seemingly correctly, that Ankita disagrees with her earlier self, and that it would make sense for her to retract her earlier assertion and to say that it is false. By contrast, contextualism seems to imply the opposite.

Unfortunately for the invariantist, her explanation of these phenomena has two problematic features. First, this explanation is universal, in the sense that it applies to all epistemic sentences. Thus, the invariantist who offers this kind of explanation of disagreement cannot concede that there is *no* disagreement when two speakers assertively utter the negations of one another's *reversible* sentences (e.g. 'water might be, and might not be, and element'), and yet maintain that there *is* disagreement when two speakers assertively utter the negations of one another's *irreversible* sentences (e.g., 'water might be an element'). For the explanation the invariantist offers for why there is disagreement in cases involving irreversible sentences will apply equally to reversible sentences.

The second and more seriously problematic feature of the invariantist explanation of disagreement and related phenomena is that it is symmetric, in the sense that it does not discriminate between situations of the better-informed speaker and the less-informed speaker. The invariantist explanation of why, in the evening, Ankita disagrees with her earlier self, and of why it would make sense for her to retract her earlier

assertion and regard it as false, is that she now sincerely asserts, and hence believes, a contradictory proposition – and, it might be added, because she now undertakes commitments that are incompatible with those she undertook in making her original assertion. But this explanation predicts, counterintuitively, that in the morning Ankita disagrees with her later self and that she should regard her later assertion as false. Similarly, this explanation predicts that Ankita should retract her earlier assertion even when her earlier self was better informed than her present self. Suppose, for example, that in the morning Ankita is the only one who knows whether Axeworthy is the murderer, she forgets this information, though she remembers that she had this information in the morning. And suppose that, in the morning, she sincerely assertively utters ~*m*, while in the evening she sincerely assertively utters *m*. The proponent of the invariantist explanations of retraction is committed to the implausible conclusion that, in the evening, Ankita should retract her earlier assertion, on the ground that she is now in a position to assert its negation.

In light of these considerations, it is natural to conclude that, contrary to the main arguments for invariantism, the latter view does not in fact do better than contextualism in accounting for disagreement and related phenomena. For while it may be argued that the contextualist *underpredicts* cases of disagreement and related phenomena, the invariantist seems to *overpredict* these phenomena, and these overpredictions seem at best to be no less problematic than the contextualist's putative underpredictions.

To rebut this charge, the invariantist would need to predict disagreement and related phenomena in cases where the contextualist counterintuitively predicts their absence, without counterintuitively predicting their presence in cases where the contextualist seems to correctly predict their absence. In the penultimate section, we will explore this possibility, and argue that it cannot succeed.

4.3 The Relativist Countermove and Why it Fails

Let us characterize, in formal terms, the dialectical situation just described, specifically as it relates to the debate between the contextualism and the relativist invariantism with respect to epistemic modals. One apparent advantage that the relativist invariantist has over the contextualist is that she can maintain the following:

> ***Knowing Assessor Principle***:[34] For any proposition, *p*, and any contexts, C_1 and C_2, if *p* is compatible with what is known in C_1 but incompatible with what is known in C_2, then an assertive utterance of *Might(p)* made in C_1 is false as assessed from C_2.

---

[34] We borrow the expressions 'knowing assessor' and 'ignorant assessor' from Dietz (2008).

where *Might*(*p*) denotes the sentence one obtains by applying an epistemic possibility modal to a sentence expressing *p*. Thus, the relativist invariantist can maintain that, as evaluated from Cavendish's context, Thales speaks falsely when he assertively utters *e* ('water might be, and might not be, an element.') On the other hand, an apparent advantage of contextualism over relativist invariantism is that the contextualist can maintain the following:

> *Ignorant Assessor Principle*: For any proposition, *p*, and any contexts, C_1 and C_2, if *p* is compatible with what is known in C_1 but incompatible with what is known in C_2, then an assertive utterance of ~*Might*(*p*) made in C_2 is true as assessed from C_1.

Thus, the contextualist can maintain that, as evaluated from Thales' perspective, Cavendish speaks truly when he assertively utters ~*e*.

It may seem, therefore, that the contextualist and the invariantist have reached a stalemate. But if there were some way in which the relativist invariantist could revise her view so as to make it compatible with both the Knowing Assessor Principle and the Ignorant Assessor Principle, then perhaps she could claim victory. MacFarlane (2011) has proposed such a revision in response to Dietz (2008).[35] On MacFarlane's proposed account, an assertive utterance of *Might*(*p*) is true, relative to a given context of assessment, just in case *p* is compatible with all the evidence that is *relevant* in the context of assessment. And the evidence that is *relevant*, in a given context of assessment, when evaluating a given assertive utterance, needn't be limited to the information that is *available* in that context of assessment; it may also include the evidence that is available in the context of the utterance being assessed. Precisely what evidence counts as relevant in a given context of assessment may depend on the interests that prevail in that context. However, MacFarlane says "in contexts where the primary point of the assessment is the critical evaluation of the speaker's assertion [let's call these *critical contexts*] … the relevant information state will generally be a composite of the speaker's and the assessor's information."[36]

MacFarlane's revised relativist view can explain both the Knowing Assessor Principle and the Ignorant Assessor Principle, on the assumption that the contexts of assessment under consideration are critical contexts. For MacFarlane's view implies that an assertive utterance of *Might*(*p*), made in one context,

---

[35] MacFarlane (2011), pp. 176-77.

[36] Note that this revised relativist account is not, strictly speaking, invariantist as we have defined this term. For, on our definition, an invariantist view is one on which the context in which an epistemic sentence is assertively uttered affects neither the content nor the truth-conditions of what is semantically expressed. However, on MacFarlane's revised relativist view, the context of utterance can affect the truth conditions of what is expressed, relative to a given context of assessment, since it can affect what information counts as relevant in this context of assessment. Nonetheless, MacFarlane's revised relativist view is similar to invariantism in that it implies that the *content* of what is asserted is invariant across contexts of utterance.

is true as assessed from a given critical context of assessment just in case p is compatible with the relevant body of evidence, one that includes both what is known in the context of utterance and what is known in the context of assessment. And if p is compatible with what is known in C_1 but incompatible with what is known in C_2, then p is incompatible with any body of evidence that includes both what is known in C_1 and what is known in C_2. And so assuming the contexts of assessment are critical contexts, MacFarlane's view implies that an assertive utterance of $Might(p)$ made in C_1 is false as assessed from C_2 (as the Knowing Assessor Principle states), and that an assertive utterance of $\sim Might(p)$ made in C_2 is true as assessed from C_1 (as the Ignorant Assessor Principle states).

Consequently, while MacFarlane does not himself propose an asymmetric account of disagreement, his revised relativism would allow for such an account. For, assuming Thales knows that Cavendish knows whether water is an element, MacFarlane's account implies that when Thales assertively utters sentence e and Cavendish assertively utters $\sim e$, Cavendish should regard Thales' assertion as false, whereas Thales should regard Cavendish's assertion as true. Hence, his account would allow us to say that, if both parties are rational, then Cavendish will disagree with Thales but Thales will not disagree with Cavendish.

Unfortunately, however, MacFarlane's view has unacceptable implications, as can be seen from the following case. Suppose Pozzo is in Potsdam and Lucky is in London. Suppose Pozzo and Lucky both recognize that Pozzo knows the temperature in Potsdam but not the temperature in London, and that Lucky knows the temperature in London but not in Potsdam. Suppose that, as a matter of fact, it's 25°C in both cities. And suppose Pozzo says:

D1. It might be hotter in London than in Potsdam.

On the view under consideration, Pozzo's assertive utterance of D1 is true, as assessed from Lucky's context, just in case it's being hotter in London than in Potsdam is compatible with the body of evidence that is relevant to Lucky's evaluation. And, assuming that Lucky is in a critical context, this body of evidence will include both what Lucky knows and what Pozzo knows. This body of evidence, therefore, entails that it's 25°C in both cities, and so it is incompatible with its being hotter in London than in Potsdam. And so MacFarlane's view has the counterintuitive implication that Pozzo's assertive utterance of D1 is false as assessed from Lucky's context.

One might think that one could solve this problem by adopting some alternative conception of what evidence counts as relevant in evaluating a speaker's assertion. However, no such maneuver can succeed without undermining the motivations for the revised relativist position. For, given some very

plausible assumptions, *any* view that is consistent with the Knowing Assessor Principle and the Ignorant Assessor Principle will imply the implausible consequence that, as evaluated from Lucky's context, Pozzo's assertive utterance of D1 is false. The argument requires two assumption. The first assumption we need is that the epistemic 'must' (understood as the dual of the epistemic 'might') is closed under conjunction. Formally

Closure under Conjunction (CC): as evaluated from a given context of assessment, if assertive utterances of *Must*(*p*) and *Must*(*q*) made in a given context of utterance would both be true, then an assertive utterance of *Must*(*p* & *q*) made in this same context of utterance would likewise be true.

And the second assumption we need is that the epistemic 'must' is closed under direct and obvious logical entailment. Formally,

> *Single Premise Closure* (SPC): as evaluated from a given context of assessment, if an assertive utterances of *Must*(*p*) made in a given context of utterance would be true, and if *p* directly and obviously entails *q*, then an assertive utterance of *Must*(*q*) made in this same context of utterance would likewise be true.

We can now derive the falsity of D1 as follows. Suppose Pozzo assertively utters the following two sentences:

> D2. It might be more than 25°C in London.

> D3. It is not the case that it might be less than 25°C in Potsdam.

By the Knowing Assessor Principle, Pozzo's assertive utterance of D2 is false as assessed from Lucky's context. And from the Ignorant Assessor Principle, Pozzo's assertive utterance of D3 is true as assessed from Lucky's context. But if, as assessed from Lucky's context, Pozzo's assertive utterance of D2 is *false*, then his assertive utterance of D2's negation, namely D4, must be true:

> D4. It must be no more than 25°C in London.

And if, assessed from Lucky's context, Pozzo's assertive utterance of D3 is *true*, then his assertive utterance of the equivalent sentence D5 must likewise be true:

> D5. It must be no less than 25°C in Potsdam.

And so it follows by CC that, assessed from Lucky's context, Pozzo's assertive utterance of the following sentence would be true.

D6. It must be no more than 25°C in London and no less than 25°C in Potsdam.

And so it follows by SPC that, assessed from Lucky's context, Pozzo's assertive utterance of the following sentence would likewise be true.

   D7. It must not be hotter in London than in Potsdam.

And since D7 is the negation of DI, we get the unacceptable result that, as assessed from Lucky's context, Pozzo's assertive utterance of DI would be false. Thus, this unacceptable result follows not only from MacFarlane's revised relativist view, but from any view that combines the Knowing Assessor Principle with the Ignorant Assessor Principle. And so this kind of compromise position, between the standard contextualist position and the standard relativist invariantist position, should not be regarded as a step forward.

Summary and Conclusion

In this paper, we have been employing reversibility cases, such as *Old News*, to shed light on the debate between contextualism and invariantism. On the contextualist view, when Ankita assertively utters *m* ('Axeworthy might be, and might not be, the murderer') in the morning, and then assertively utters ~*m* in the evening, there is no conflict between the two propositions she asserts, and so this view seems to imply that in the evening she does not disagree with her earlier self. The invariantist aims to avoid this kind of implication by proposing a view on which there is genuine conflict between these assertions, and hence genuine disagreement.

   We have presented two problems for the invariantist view. First, as we argued in parts 2 and 3, this view cannot explain how Ankita can genuinely disagree with her earlier self in the evening without at the same time predicting that it is irrational for her to sincerely assertively utter *m* in the morning. For the versions of invariantism that can explain genuine disagreement (absolutist invariantism, and versions of relativist invariantism that posit pragmatic conflict) imply that Ankita is irrational in making her sincere assertion in the morning, whereas versions of invariantism that are consistent with Ankita's being rational in the morning (versions of relativist invariantism that do not posit pragmatic conflict) fail to explain any genuine disagreement.

The second, related problem we raised for the invariantist view is that its account of disagreement is symmetric: just as it predicts that in the evening Ankita disagrees with her earlier self, so likewise it predicts that in the morning she disagrees with her future self. And as we argued in part 4, the latter implication is implausible. We noted that the invariantist can avoid this implication by adopting a revised form of relativism according to which, while Ankita's morning assertion is false relative to her evening context, her evening assertion is true relative to her morning context. But this revised view, we have shown, has its own unpalatable consequences.

It seems, therefore, that there are serious problems with the standard relativist view, according to which the epistemic assertions of better-informed speakers are false as assessed from the contexts of less-informed speakers. And there are also serious problems with the revised relativist view, according to which the assertions of less-informed speakers are false as assessed from the contexts of better-informed speakers, but not vice versa. And so the contextualist view — on which neither the assertions of the better-informed speakers, nor the assertions of the less-informed speaker, are false as assessed from the context of the other — is more plausible than it may first have appeared.

Does this mean we should accept contextualism? Not necessarily. For there remains the burden of explaining why it is that when a less-informed and a better-informed speaker assert the negations of one another's epistemic sentences, the better-informed speaker *at least appears* to disagree with the less-informed speaker. We have shown that we should not accept the invariantist explanation of this phenomenon. But what explanation we should accept — whether, for example, we should accept an explanation of the kind proposed by sophisticated versions of contextualism, or instead an explanation that draws on the resources of non-propositional views — remains to be seen.[37, 38]

---

[37] For discussions of how contextualists can account for disagreement, see Dietz (2008) and Björnson and Finlay (2010). For alternative views on disagreement, see von Fintel and Gillies (2007) and (2008). For non-propositional theories of the epistemic expressions, see particularly Veltman (1996), Gillies (2004), and Yalcin (2011).

[38] This paper is the product of five years of collaboration, and gratitude is due to more people than we can remember. Of those we can, special thanks to John MacFarlane, Andy Egan, Scott Soames, Shieva Kleinschmidt, Julia Staffel, Johannes Schmitt, Ben Lennertz, Mike Titelbaum, and anonymous referees for *Mind* for invaluable comments and discussion.

References

Arntzenius, Frank 2003: 'Some Problems for Conditionalization and Reflection.' *Journal of Philosophy*, 100, pp. 356-70.

Bach, Kent 2011: 'Perspectives on Possibility: Contextualism, Relativism, or What.' In Egan and Weatherson 2011, pp. 19-59.

Björnson and Finlay 2010: 'Metaethical Contextualism Defended.' *Ethics*, 121, pp. 7-36.

Bovens, Luc 1995: '"P and I will believe Not-P": Diachronic Constraints on Rational Belief .' *Mind*, 104, pp. 737-60.

Briggs, Rachael 2009: 'Distorted Reflection.' *Philosophical Review*, 118, pp. 59-85.

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, pp. 225-54.

Christensen, David 1991: 'Clever Bookies and Coherent Beliefs.' *Philosophical Review*, 100, pp. 229-47.

Dietz, Richard 2008: 'Epistemic Modals and Correct Disagreement.' In García-Carpintero and Kölbel 2008, 239-64.

Egan, Andy 2007: 'Epistemic Modals, Relativism, and Assertion.' *Philosophical Studies*, 133, pp. 1-22.

_____ 2010: 'Disputing About Taste.' In Feldman and Warfield 2010, pp. 247-86.

Egan, Andy, John Hawthorne, and Brian Weatherson 2005: 'Epistemic Modals in Context.' in Preyer and Peter 2005, pp. 131-68.

Egan, Andy, and Brian Weatherson (eds) 2011. *Epistemic Modality*. Oxford: Oxford University Press.

Elga, Adam 2000: 'Self-locating Belief and the Sleeping Beauty Problem.' *Analysis*, 60, pp. 143-7.

Feldman, Richard, and Ted Warfield (eds) 2010: *Disagreement*. Oxford: Oxford University Press.

von Fintel, Kai, and Anthony Gillies 2007: 'An Opinionated Guide to Epistemic Modality.' *Oxford Studies in Epistemology*, 2, pp. 32-62.

von Fintel, Kai, and Anthony Gillies 2008: 'CIA Leaks.' *Philosophical Review*, 117, pp. 77-98.

García-Carpintero, Manuel, and Max Kölbel (eds) 2008: *Relative Truth*. Oxford: Oxford University Press.

Gibbard, Allan 1981: 'Two Recent Theories of Conditionals.' In Harper, Stalnaker, and Pearce 1981, pp. 211-48.

_____ 1990: *Wise Choices, Apt Feelings*. Cambridge, MA: Harvard University Press.

Gillies, Anthony 2004: 'Epistemic Conditionals and Conditional Epistemics.' *Noûs*, 38, pp. 585-616.

Harper, William, Robert Stalnaker, and Glenn Pearce (eds) 1981: *Ifs: Conditionals, Beliefs, Decisions, Chance, and Time*. Dordrecht: Reidel.

Kölbel, Max 2002: *Truth Without Objectivity*. New York: Routledge.

Kolodny, Niko, and John MacFarlane 2010: 'Ifs and Oughts.' *Journal of Philosophy*, 107, pp. 115-43.

Lasersohn, Peter 2005: 'Context Dependence, Disagreement, and Predicates of Personal Taste.' *Linguistics and Philosophy*, 28, pp. 643-86.

Lewis, David 1979: 'Attitudes *De Dicto* and *De Se*.' *Philosophical Review*, 88, pp. 513-43.

MacFarlane, John 2003: 'Future Contingents and Relative Truth.' *Philosophical Quarterly*, 53, pp. 321-36.

_____ 2005: 'Making Sense of Relativist Truth.' *Proceedings of the Aristotelian Society*, 105, pp. 321-39.

_____ 2007: 'Relativism and Disagreement.' *Philosophical Studies*, 132, pp. 17-31.

_____ 2011: 'Epistemic Modals are Assessment-Sensitive.' In Egan and Weatherson 2011, pp. 144-78.

_____ forthcoming: 'Richard on Truth and Commitment.' Forthcoming in *Philosophical Studies*.

_____ 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 11/9/11, available at http://johnmacfarlane.net/reltruth-chapters.pdf.

Parfit, Derek 2011: *On What Matters*, volumes I and II. Oxford: Oxford University Press.

Perry, John 1979: 'The Essential Indexical.' *Noûs*, 13, pp. 3-21.

Preyer, Gerhard, and Georg Peter (eds) 2005: *Contextualism in Philosophy*. Oxford: Oxford University Press.

Price, Huw 1983: 'Does 'Probably' Modify Sense?' *Australasian Journal of Philosophy*, 61, pp. 396-408.

Regan, Donald 1980: *Utilitarianism and Cooperation*. Oxford: Oxford University Press.

Richard, Mark 2004: 'Contextualism and Relativism.' *Philosophical Studies*, 119, pp. 215-42.

Ross, Jacob 2006: *Acceptance and Practical Reason*. Doctoral Dissertation, Rutgers University.

Schroeder, Mark 2008; *Being For: Evaluating the Semantic Program of Expressivism*. Oxford: Oxford University Press.

_____ forthcoming: 'Two Roles for Propositions: Cause for Divorce?' Forthcoming in *Noûs*.

_____ unpublished: 'Attitudes and Epistemics.' Unpublished paper. Available at http://www-bcf.usc.edu/~maschroe/research/Schroeder_Attitudes_and_Epistemics.pdf.

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, pp. 487-525.

_____ 2007 b; 'Indicative Conditionals Have Relative Truth Conditions.' *Proceedings of the Annual Meeting of the Chicago Linguistics Society*, 43, pp. 231-42.

Stevenson, C. L. 1937: 'The Emotive Meaning of Ethical Terms'. Reprinted in his 1963, 10-31.

_____ 1963. *Facts and Values*. Westport, CT: Greenwood Press.

Talbot, William 1991: 'Two Principles of Bayesian Epistemology.' *Philosophical Studies*, 62, pp. 135-50.

van Fraassen, Bas 1984: 'Belief and the Will'. *Journal of Philosophy*, 81, pp. 235-56.

Veltman, Frank 1996: 'Defaults in Update Semantics.' *Journal of Philosophical Logic*, 25, pp. 221-61.

Weatherson, Brian 2009: 'Conditionals and Indexical Relativism.' *Synthese*, 166, pp. 333-57.

Wright, Crispin 2007: 'New Age Relativism and Epistemic Possibility: The Question of Evidence.' *Philosophical Issues*, 17, pp. 262-83.

Yalcin, Seth 2007: 'Epistemic Modals.' *Mind*, 116, pp. 983-1026.

_____ 2011: 'Nonfactualism About Epistemic Modality.' In Egan and Weatherson 2011, pp. 295-332.

Yalcin, Seth and Josua Knobe 2007: 'Fat Tony Might Be Dead: An Experimental Note on Epistemic Modals.' *Semantics Archive* 10-12-30.