

# SUBJECTIVE DISAGREEMENT

Bob Beddar

forthcoming in *Noûs*

## 1 Introduction

People often disagree over objective matters. For example, Ava believes that global warming has human causes; Bert believes it doesn't. There is a natural story to be told about such objective disagreements. Disagreement over some objective matter, so the story goes, is disagreement about what the world is like. This natural story comes with a natural formal implementation. Suppose we adopt the familiar model of a proposition as a set of possible worlds. Then we can say that A and B have an objective disagreement over some proposition  $p$  if and only if A believes  $p$  and B believes  $\neg p$  (or *vice versa*). Thus Ava and Bert disagree over the set of worlds in which global warming has human causes—she believes it, and he disbelieves it.

But people also disagree over subjective matters.<sup>1</sup> To start with an example that has featured prominently in the recent literature, consider disagreements over matters of taste.<sup>2</sup> Ava and Bert share a plate of escargot: she finds it tasty; he thinks it's disgusting. In so believing, it seems that they disagree. But are they disagreeing over any objective matter? This seems doubtful. Presumably, the world doesn't contain any facts about what is tasty (or disgusting, or funny etc.) *simpliciter*: it only contains facts about what is tasty (or disgusting, or funny, etc.) to specific agents. But it seems Ava and Bert could agree about all of *these* facts—they could agree that the escargot is tasty to her, and disgusting to him—while still disagreeing over whether the escargot is tasty.

---

<sup>1</sup>What I call, ‘subjective disagreement’ is often discussed under the label, ‘faultless disagreement’ (see e.g., Kölbel 2003). However, I think the notion of faultlessness is open to multiple interpretations, and is liable to mislead. In particular, ‘faultlessness’ is naturally understood as entailing a lack of criticizability. This sort of faultlessness can come apart from subjectivity: in principle, a disagreement could be subjective (in the sense that it does not latch onto any objective facts) even though one party is criticizable for their beliefs. (On some views, moral disagreements are like this.)

<sup>2</sup>For relevant discussion, see Lasersohn 2005; Stephenson 2007a,c; Stojanovic 2007; Cappelen and Hawthorne 2009; Egan 2010, 2014; Huvenes 2012; Schaffer 2011; Sundell 2011; Barker 2013; Pearson 2013; MacFarlane 2014; Marques and García-Carpintero 2014, among many others.

Next, consider disagreements involving the “language of subjective uncertainty” (Swanson 2006)—broadly construed to include epistemic modals (*might*, *must*), probability operators (*probably*, *likely*), and indicative conditionals. On the basis of the initial crime scene canvas, the lieutenant believes the butler might be the culprit. However, the detective has just checked the butler’s alibi; as a result, she thinks there’s no chance the butler could have done it. As a number of authors have noted, there is some inclination to regard them as disagreeing—they disagree over whether the butler might be guilty.<sup>3</sup> But here too it is far from clear that this is a disagreement over what the world is like.

There are various other candidates for subjective disagreements. Take, for example, disagreements over aesthetic or moral matters. According to a variety of antirealist views, aesthetic and moral matters are much like matters of taste: just as there are no objective facts about what is tasty, there are no objective facts about what is beautiful or what is morally permissible. And some authors have extended this antirealist view to all normative matters.<sup>4</sup> According to these authors, disagreements about, say, rationality and justification are also importantly different from objective disagreements.

Much work in recent philosophy of language and semantics has been driven by a desire to make sense of such subjective disagreements. For example, relativists about some class of subjective expressions<sup>5</sup> often object to contextualism on the grounds that disagreement requires *shared content*: if A disagrees with B, there must be some content that A accepts and B rejects. This in turn is used to motivate the idea that certain contents have relative truth-values: they can be true relative to one assessor, and false relative to another.<sup>6</sup> Expressivists usually take a different tack, suggesting that many subjective disagreements should be understood in terms of “disagreement in attitude”—a conflict in the relevant parties’ desire-like states.<sup>7</sup>

While many participants in these debates make assumptions about the conditions under which disagreements occur (or fail to occur), surprisingly few have ventured a general theory of subjective disagreement. And the few theories that have been offered face serious difficulties. In particular, extant theories either predict *too much* subjective disagreement or *too little*. In this paper, I develop an account of subjective disagreement

---

<sup>3</sup>See e.g., Egan et al. 2005; Egan 2007; Stephenson 2007a,b,c; MacFarlane 2011; Willer 2013. Others have questioned these judgments—see e.g., Yalcin 2011; Knobe and Yalcin 2014.

<sup>4</sup>See e.g., Gibbard 1990, 2003; Field 2009.

<sup>5</sup>By a ‘subjective expression’, I mean any expression that gives rise to subjective disagreement—e.g., taste predicates, epistemic modals, etc. (Ross and Schroeder (2013) use the term, ‘contested expression’; as far as I am concerned, the two labels are notational variants.)

<sup>6</sup>For versions of this relativist argument, see Richard 2004, 2008; Lasersohn 2005; Stephenson 2007a,c; Egan 2010, 2012, 2014; MacFarlane 2011, 2014. For responses, see Glanzberg 2007; de Sa 2008; Cappelen and Hawthorne 2009; Schaffer 2011; Sundell 2011; Huvenes 2012; Pearson 2013; Plunkett and Sundell 2013; among others.

<sup>7</sup>See e.g., Stevenson 1944, 1963; Blackburn 1984; Gibbard 1990, 2003.

that fares better. I then consider the implications of my theory for the semantics of subjective expressions.

The first half of the paper is critical: I examine current treatments of subjective disagreement and explain why they are extensionally inadequate. I start (§2) by considering the natural idea that subjective disagreement arises from incompatibility in *de se* contents. This proposal predicts too much subjective disagreement: in particular, it predicts that when I have a *de se* belief that I'm hungry, and you have a *de se* belief that you're full, we count as disagreeing. §3 considers whether we can understand subjective disagreement as a discursive phenomenon—one that arises when interlocutors propose incompatible updates to the common ground. This proposal predicts too little disagreement: in particular, it fails to explain how agents can disagree over subjective matters even when they are not conversing. §4 examines whether subjective disagreement can be understood in terms of 'disagreement in attitude'. This proposal also predicts either too much disagreement or too little, depending on how one cashes out the notion of disagreement of attitude.

After the negative comes the positive: the second half of the paper develops a novel strategy for understanding disagreement. The strategy is to semantically ascend and focus on analyzing *disagreement ascriptions*. §5 offers linguistic data suggesting that our intuitions about the truth or falsity of disagreement ascriptions are not determined solely by the contents of the relevant agents' propositional attitudes; rather, they are sensitive to syntactic features of the sentences we use to characterize these attitudes. I go on (§6) to leverage this observation into an analysis of disagreement ascriptions in terms of inconsistency relations that obtain between the sentences used to characterize agents' beliefs. I show that this analysis delivers neither too much nor too little subjective disagreement. Moreover, it can be reconciled with both relativist and contextualist treatments of subjective expressions. One important upshot of this analysis is that appeals to subjective disagreement won't settle the debate between contextualists and relativists; rather, the battle must be fought on other grounds. §7 concludes by discussing how subjective disagreement relates to objective disagreement on my account.

## 2 Subjective Disagreement as *De Se* Conflict

We started with the attractive idea that objective disagreement arises when one party believes a possible worlds proposition and another party believes its negation. But it is controversial whether the contents of all beliefs can be modeled by sets of worlds. In particular, the contents of *de se* beliefs (that is, beliefs about one's location within the world) are notoriously difficult to model using just sets of worlds. In response to this difficulty, one common approach is to follow Lewis (1979) in modeling the contents of *de*

*se* beliefs using sets of centered worlds, where a centered world is an ordered pair of a world and agent inhabiting that world.<sup>8,9</sup>

This suggests a natural hypothesis. Perhaps subjective disagreement arises when two parties have inconsistent *de se* beliefs. In other words:

**De Se Conflict** Subjective disagreement occurs when one party believes a centered worlds proposition  $p$ , and another party believes its negation.

To illustrate, consider again the gustatory dispute between Ava and Bert. According to *De Se Conflict*, for Ava to believe that the escargot is tasty is for her to believe a centered proposition such as:

$$\text{TASTY} = \{\langle w, i \rangle \mid i \text{ enjoys the way the escargot tastes at } w\}.$$

For Bert to believe that the escargot isn't tasty is for him to believe the negation of *TASTY* (i.e., the set of  $w, i$  pairs such that  $i$  doesn't enjoy the way the escargot tastes at  $w$ ). Hence, by *De Se Conflict*, they disagree. Other candidate subjective disagreements—e.g., disagreements about what might be the case, or what ought to be the case—will be handled in similar fashion: for each such disagreement, the strategy will be to regard the parties as believing inconsistent centered propositions.

The hypothesis that subjective disagreement arises from *de se* conflict is attractive for at least two reasons. First, it really does seem that many of the subject matters that give rise to subjective disagreements have a *de se* dimension. This is perhaps clearest when it comes to matters of taste.<sup>10</sup> Consider: what does it take for a belief report such as (1) to be true?

(1) Ava believes [/thinks] the escargot is tasty.

(1) does not seem to require that Ava believes the escargot is tasty to those around her. After all, (1) could be true even if she knows that her dining companion, Bert, loathes the taste of escargot. Similarly, (1) does not seem to require that Ava believes that escargot is tasty to *most* people, or to a generic individual—she might have no view on the matter; she might even acknowledge that her tastes are idiosyncratic. Indeed, (1) does not even seem to require that Ava has a *de dicto* belief of the form: *The escargot is tasty for Ava*. After

---

<sup>8</sup>Some authors model centered worlds as ordered triples of worlds, times, and agents. For my purposes, I will set this complication aside.

<sup>9</sup>Of course, the Lewisian account of the *de se* is not without challengers. For a rival approach, see Perry 1979. For skepticism about whether the *de se* deserves special semantic treatment, see Cappelen and Dever 2013. While my sympathies lie with the Lewisian account, I will not try to argue for it here.

<sup>10</sup>The idea that taste beliefs are *de se* is closely associated with relativism, and plays a central role in the relativist proposals of Stephenson (2007a,c) and Egan (2010). However, it is at least open to contextualists to agree that taste beliefs are *de se* (a point made by Pearson (2013)). In §6, I show in detail how both relativists and contextualists can model the contents of taste beliefs using centered propositions.

all, (1) could be true even if Ava is an amnesiac, and lacks any *de dicto* beliefs about Ava's gastronomical preferences.

At the same time, (1) seems to require more than just a *de re* belief that the escargot is tasty to some *res*, where that *res* turns out to be Ava (Stephenson 2010). To see this, consider a scenario in which Ava despises the taste of the escargot (and knows that she does). However, Ava glimpses a woman across the restaurant eating the molluscs, and comes to believe: *That woman enjoys the taste of the escargot*. Unbeknownst to Ava, the woman is none other than Ava herself; she caught her own reflection in a mirror. Intuitively, (1) is false in this situation. This suggests that (1) is naturally read as ascribing to Ava a *de se* belief that she herself enjoys the taste of the escargot.<sup>11</sup>

Arguably, we also find this *de se* dimension in beliefs involving epistemic modals (Egan 2007; Stephenson 2007a,c). Consider: what does it take for (2) to be true?

(2) The lieutenant believes [/thinks] the butler might be guilty.

(2) does not seem to require that the lieutenant believes that the butler's guilt is compatible with what some *other* folks know. After all, (2) could be true even if the lieutenant is aware that other investigators have gathered additional evidence—evidence that may exonerate the butler. At the same time, (2) seems to require more than a mere *de re* belief that the butler's guilt is compatible with the knowledge of some *res*, where that *res* turns out to be the lieutenant. To see this, consider a scenario in which the lieutenant is convinced of the butler's innocence, but thinks that the butler's guilt is compatible with what *that person over there* knows, where *that person over there* turns out to be the lieutenant. Intuitively, (2) is false in this scenario. This suggests that (2) is naturally read as ascribing a *de se* belief. On this reading, (2) claims that the lieutenant believes a centered proposition such as:

$$\text{GUILTY} = \{\langle w, i \rangle \mid \text{it's compatible with what } i \text{ knows that the butler is guilty at } w\}.$$

---

<sup>11</sup>An important qualification is in order: as a number of writers have noted, we can sometimes access 'exocentric' readings of taste predicates in attitude reports—that is, readings where the taste predicate makes reference to the taste of some individual other than the attitude holder (Lasersohn 2005; Stephenson 2007a,c; Cappelen and Hawthorne 2009). For example, suppose Ava watches her dog Fido consume the new puppy chow with an unusual degree of relish. Then it seems we can truthfully say:

(i) Ava thinks the puppy chow is tasty.

meaning that she thinks it's tasty to Fido.

Does the availability of such readings undermine the idea that taste beliefs are *de se*? While I agree that (i) has a reading that does not ascribe to Ava a belief in TASTY, it also seems that we can access a different reading of (i) on which it does. That is, we can access a reading of (i) on which it claims that Ava believes *de se* that she enjoys the taste of the puppy chow. This shows that attitude reports embedding taste predicates typically have *de se* readings. And our reaction to sentences such as (1) suggests that this reading is the default; it is only overridden in (i) by our awareness that humans usually dislike the taste of puppy chow.

Thus *De Se* Conflict captures the insight that many beliefs that give rise to subjective disagreements have a *de se* dimension.<sup>12</sup> A second attractive feature of *De Se* Conflict is that allows for a simple, unified account of all disagreement, both objective and subjective:

**Simple Account** A and B disagree over  $p$  if and only if A believes  $p$  and B believes  $\neg p$ .<sup>13</sup>

According to this proposal, subjective disagreements arise when  $p$  is a centered proposition. Objective disagreements arise when  $p$  is a traditional possible worlds proposition, or a ‘boring’ centered proposition (that is, a centered proposition that is equivalent to a set of worlds).<sup>14</sup>

Despite these attractions, the hypothesis that subjective disagreement arises from conflicts in *de se* content faces a fatal flaw. As various writers have observed, it overgenerates disagreements among run-of-the-mill *de se* beliefs.<sup>15</sup> An example: Kim is hungry, and she knows it. Sara has just eaten; consequently, she knows that she (Sara) isn’t hungry. According to the Lewisian account of *de se*, the content of Kim’s belief is the set:

$$\text{HUNGRY} = \{\langle w, i \rangle \mid i \text{ is hungry at } w\}.$$

The content of Sara’s belief is the negation of HUNGRY (i.e., the set of  $w, i$  pairs such that  $i$  is not hungry at  $w$ ). Given *De Se* Conflict, it follows that they disagree. But this is surely wrong.<sup>16</sup>

Some might suggest that this difficulty can be avoided by replacing *De Se* Conflict with a close variant. While Kim and Sara’s beliefs have inconsistent contents, their beliefs can both be accurate, in the sense that Kim’s belief can be true of *Kim* and Sara’s belief can be true of *Sara* at the same world. And so, one might suggest, disagreement is not a matter of having inconsistent beliefs; rather, it’s a matter of having beliefs that cannot both be accurate (MacFarlane 2007):

---

<sup>12</sup>Arguably, beliefs about other subjective matters also have a *de se* dimension. For example, see Lewis (1989) and Egan (2012) for the view that value judgments are *de se*.

<sup>13</sup>As MacFarlane (2014) notes, the Simple Account is probably the default story about disagreement for most philosophers. For defenses of the Simple Account, see Kölbel 2003; Cappelen and Hawthorne 2009; Lasersohn 2013.

<sup>14</sup>More precisely, say that a centered proposition  $p$  is boring iff for any individuals A, B, and any world  $w$ ,  $\langle w, A \rangle \in p$  iff  $\langle w, B \rangle \in p$  (Egan 2006: 107; Ninan 2010: 553).

<sup>15</sup>See, in particular, Dreier 2009 for a forceful statement of this objection. Versions of this problem are also discussed in Egan 2010, 2012, 2014; Richard 2011, and Ross and Schroeder 2013.

<sup>16</sup>As MacFarlane (2007, 2014) observes, a similar difficulty arises for anyone who believes in temporal propositions. Suppose that at noon Fred believes the temporal proposition, *(It’s sunny outside)*; with the fall of dusk, Ted comes to believe the temporal proposition, *(It’s not sunny outside)*. The Simple Account incorrectly predicts that midday Fred disagrees with crepuscular Ted over the proposition, *(It’s sunny outside)*.

**No Joint Accuracy** Suppose A believes  $p$  and B believes  $q$ . These beliefs disagree iff they cannot both be accurate—i.e.,  $\nexists w$  s.t.  $\langle w, A \rangle \in p$  and  $\langle w, B \rangle \in q$ .<sup>17</sup>

But while this avoids overgenerating disagreement among ordinary *de se* beliefs, it does so at the cost of undergenerating subjective disagreement. After all, if taste belief is *de se*, then Ava’s belief that the escargot is tasty and Bert’s belief that the escargot isn’t tasty could both be accurate: all that’s required is for Ava to enjoy the way the escargot tastes and for Bert to fail to do so. Similar remarks apply to other candidate subjective disagreements (e.g., those involving epistemic modals, those involving moral matters).<sup>18</sup>

To take stock: appealing as it may seem, we cannot explain subjective disagreement in terms of the contents of *de se* beliefs. *De Se* Conflict saddles us with too much subjective disagreement, and No Joint Accuracy leaves us with too little.<sup>19</sup>

### 3 Subjective Disagreement as Discursive Disagreement

A different approach to subjective disagreement is to focus on *discursive disagreement*—that is, disagreements that arise in discourse. For example, we can imagine Ava and Bert engaging in the following exchange:

- (3) a. *Ava*: This is tasty!  
b. *Bert*: No [/Nuh uh], it’s disgusting!

Indeed, the majority of work on disagreements involving subjective expressions has focused on such discursive disagreements. Perhaps, then, a proper analysis of discursive disagreement will yield an adequate account of subjective disagreement in general.

---

<sup>17</sup>Note that whenever  $p$  and  $q$  are boring centered propositions the Simple Account and No Joint Accuracy will yield the same verdict about whether two beliefs disagree.

<sup>18</sup>While my primary concern is whether No Joint Accuracy gives us enough disagreement, MacFarlane (2007) also raises a worry about whether No Joint Accuracy overpredicts disagreement in cases where one person believes a necessary tensed proposition  $p$  at time  $t_1$  and another person disbelieves  $p$  at  $t_2$ . (Despite raising this difficulty, MacFarlane (2007) appears to accept No Joint Accuracy, at least as a working assumption.)

<sup>19</sup>Could we escape this dilemma by rejecting the Lewisian account of the *de se*? There are at least two difficulties with this approach. First, it’s not clear that we should make our account of *de se* content hostage to a theory of subjective disagreement (especially since the Lewisian account of the *de se* has so many explanatory virtues). Second, and more decisively, it’s doubtful that an alternative account of the *de se* will secure us enough subjective disagreement. For example, suppose one takes a *de se* belief to be a three-place relation between an agent, a possible worlds (or structured) proposition, and a first-person propositional guise or mode of presentation. Combining this proposal with the Simple Account once again undergenerates subjective disagreements. After all, the proposition that Ava believes when she believes the escargot is tasty is:  $\{w \mid \text{Ava enjoys the way the escargot tastes at } w\}$  (or its structured counterpart). And this proposition is perfectly compatible with the proposition that Bert believes when he believes that the escargot isn’t tasty ( $\{w \mid \text{Bert doesn’t enjoy the way the escargot tastes at } w\}$ , or its structured counterpart).

What would such an analysis look like? There are various ways of developing an account of discursive disagreement. In the interests of space, I will focus on an approach that has occupied a particularly prominent place in the recent literature. According to the account in question, a linguistic exchange constitutes a discursive disagreement when the interlocutors make incompatible proposals for updating the conversational context:

**Discursive Disagreement** A and B have a discursive disagreement if and only if they make utterances with incompatible uptake conditions.

An account along these lines can be implemented in either a relativist or contextualist setting. For example, [Egan \(2007, 2010, 2012, 2014\)](#) defends a relativist implementation of this idea. According to Egan, the content of Ava's assertion of (3a) is a centered proposition such as **TASTY**. In so asserting, Ava is making a conversational move whose conventional effect is to get one's audience to believe its content—i.e., to self-attribute the property of enjoying the taste of the escargot. By contrast, the content of Bert's assertion of (3b) is a centered proposition that is inconsistent with the content of Ava's utterance (e.g.,  $\neg \text{TASTY}$ ). In so asserting, Bert is making a conversational move whose conventional effect is to get one's audience to believe *its* content. Since these two assertions have incompatible uptake conditions, they constitute a discursive disagreement.<sup>20</sup>

Other authors have developed Discursive Disagreement in a contextualist framework. For example, some have proposed that Ava and Bert are involved in a metalinguistic dispute about how to set the value of a conversational standard of taste parameter ([Plunkett and Sundell 2013; Barker 2013](#); for similar ideas, see [DeRose 2004; Silk forthcoming; Khoo 2015; Khoo and Knobe forthcoming](#)). Thus in uttering (3a), Ava proposes to set the standard of taste parameter in such a way that the escargot falls under the extension of *tasty*; in uttering (3b), Bert rejects this proposal.<sup>21</sup>

Let us grant—at least for the sake of argument—that an account along these lines can handle discursive disagreements involving subjective expressions. Even if we grant this much, there remains a worry that we won't have accounted for *all* cases of subjective disagreement. In particular, we won't have accounted for *speechless disagreements*—cases in which two parties disagree over some matter, but this disagreement never gets voiced.

Such speechless disagreements occur all of the time. Certainly we have speechless disagreements concerning objective matters: if Ava believes global warming has human causes, and Bert believes it doesn't, then there's a very real sense in which they disagree, even if they never have a conversation about it.<sup>22</sup> And there also seem to be plenty of

---

<sup>20</sup>See [Stephenson 2007a](#) for a very similar account.

<sup>21</sup>For other contextualist proposals that make use of some notion of discursive disagreement, see [Glanzberg 2007; Sundell 2011; de Sa 2008, 2015; Pearson 2013](#).

<sup>22</sup>Compare [Cappelen and Hawthorne 2009](#): 60–61 on agreement as a “state” vs. an “activity”.

speechless disagreements over subjective matters. For example, we can imagine a variant of our dining scenario in which Ava and Bert are having an awkward, silent first date. Ava tries the escargot; she finds it delicious. Encouraged by the relish with which she consumes the gastropods, Bert tries a bite and finds it disgusting. Once again, it seems natural to describe them as disagreeing: they disagree about whether the escargot is tasty, even though neither vocalizes this disagreement.

Similar remarks apply to other candidate subjective disagreements. Take moral matters: if I believe that abortion is permissible whereas you believe it is impermissible, then we disagree even if we never get into a conversation about the matter. Or take epistemic modals: if the lieutenant believes the butler might have done it and the detective believes he couldn't have, then there's at least some inclination to describe them as disagreeing over whether the butler might have done it, even if they never discuss the matter. More generally, for pretty much any class of subjective disagreements, it seems we can concoct speechless disagreements in that domain. This bodes ill for attempts to reduce all subjective disagreement to discursive disagreement.

Some might suggest that we can characterize speechless disagreements in terms of the conditions under which the parties *would* have a discursive disagreement, were they to enter into a conversation. As Egan observes, "One way for my believing  $p$  and your believing  $q$  to amount to a disagreement is for it to be likely, were we to get into a conversation, to lead to [a discursive disagreement]" (2014: 96). Here is a first pass attempt to flesh this out:

**Discursive Account of Speechless Disagreement** Two parties have a speechless disagreement over some subjective matter  $M$  iff they would be likely to have a discursive disagreement, were they to discuss  $M$ .

It's easy to come up with counterexamples to this first pass account. For a counterexample to the sufficiency condition, imagine that A and B hold the exact same opinion on every matter. However, A is a contrarian who will contradict anything B says, for the sheer pleasure of it. For any subjective matter  $M$ , the biconditional wrongly predicts that A and B disagree, since they would have a discursive disagreement over  $M$  were they to start conversing. For a counterexample to the necessity condition, imagine that A and B have differing opinions on some subjective matter. But A is a sycophant who invariably assents to whatever B says. The biconditional incorrectly predicts that there's no disagreement between A and B.

It might seem that there's an easy way of fixing this first pass account: simply require that the discussion in question needs to be *sincere*. But even if we make this amendment, further counterexamples are in store. Suppose A and B speechlessly disagree about  $M$ , but

A regards B as his epistemic superior: were they to converse, A would immediately defer to B, hence no discursive disagreement would ensue. To deal with such cases, we could try further amendments; perhaps we should stipulate that neither party would revise their beliefs throughout the course of the conversation. However, this won't be enough to stem the tide of counterexamples. Suppose that A and B speechlessly disagree about *M*; however, A is *assertorically timid*: due to an unhealthy fear of asserting falsehoods, he would never make a flat-out assertion about *M*; instead, he would only venture hedged assertions (e.g. 'I think [/suspect] that...'). Our amended Discursive Account incorrectly predicts that A and B do not speechlessly disagree.

Those attracted to the Discursive Account may be inclined to press ahead, adding further conditions and qualifications. But I think we should be worried on their behalf: a steady stream of counterexamples is often a symptom that a particular theory is on the wrong track, in which case the right response is to scrap the theory, rather than tinker with the details. What's more, even if by adding further conditions we can arrive at a Discursive Account that is free from counterexamples, there's a natural worry that the resulting account will be *ad hoc* and unmotivated.

To feel the force of this worry, just consider the two conditions we've already mentioned: the requirement that the discussion needs to be sincere, and the requirement that neither party would revise their beliefs during the discussion. What work are these conditions doing? It seems that their main function is to ensure that the assertions that constitute the discursive disagreement about *M* express the contents of the parties' beliefs about *M*. This in turn suggests that what's really doing the explanatory work is the contents of the beliefs. But this is to give up the whole idea behind the Discursive Account, plunging us back into the troubled waters of explaining subjective disagreement in terms of doxastic inconsistency (§2).<sup>23</sup>

---

<sup>23</sup>So far I've formulated the Discursive Account in counterfactual terms: if such-and-such conditions were to obtain, a discursive disagreement would arise. Some may observe that counterfactual analyses have a pretty spotty track record (see e.g., Shope 1978). However, in at least some cases where counterfactual analyses fall victim to counterexample, analyses stated in terms of dispositions fare better. And so, some may suggest, we should recast the Discursive Account as a dispositional theory, e.g.:

**Discursive Account of Speechless Disagreement (Dispositional)** Two parties have a speechless disagreement over some subjective matter *M* iff they are disposed to have a discursive disagreement over *M*.

But even if recasting the Discursive Account in dispositional terms avoids some counterexamples, it certainly doesn't avoid all of them. Consider again the contrarian. It's not just that there's a true counterfactual of the form, 'If the contrarian were to have a conversation with his companion, a discursive disagreement would ensue.' It's also true that the contrarian is *disposed* to have a discursive disagreement with his companion. Thus the contrarian refutes the sufficiency condition of the dispositional Discursive Account. Similarly, the sycophant and the timid asserter refute the necessity condition: both speechlessly disagree with their companions, but neither is disposed to have discursive disagreements with them. And so couching a Discursive Account in dispositional terms does not solve the problem.

Thus it is unlikely that we'll be able to leverage an account of discursive disagreement into a fully general theory of subjective disagreement. This is not to say that there's no connection between speechless disagreement and discursive disagreement; it's certainly true that disagreements in beliefs often give rise to discursive disagreements. But, as the foregoing examples show, it proves quite difficult to move from this observation to a general analysis of subjective disagreement. If we'd like such an analysis, we had best look elsewhere.

## 4 Subjective Disagreement as Disagreement in Attitude

In the metaethics literature, there is a long tradition of insisting that conflicts in agents' desires (or desire-like attitudes) can constitute a sort of disagreement: 'disagreement in attitude'.<sup>24</sup> Stevenson gives the following example:

Suppose that two people have decided to dine together. One suggests a restaurant where there is music; another expresses his disinclination to hear music and suggests some other restaurant... The disagreement springs more from divergent preferences than from divergent beliefs, and will end when they both *wish* to go to the same place... (1944: 3)

If we can make sense of the notion of disagreement in attitude, perhaps we could use it to explain subjective disagreements.

There are a couple ways of developing this strategy. One option is to retain Discursive Disagreement and supplement it with a theory of disagreement in attitude in order to explain speechless disagreements.<sup>25</sup> To illustrate, recall Ava and Bert's speechless disagreement over whether the escargot is tasty. According to the proposal under consideration, they disagree in virtue of the fact that Ava has some desire regarding the escargot—a desire that conflicts with Bert's desires vis-à-vis the escargot.

Another option is to take disagreement in attitude to be explanatorily prior to discursive disagreement. For example, we might follow expressivists in insisting that one way for a particular linguistic exchange to constitute a discursive disagreement is for the interlocutors' utterances to express conative attitudes that disagree with one another. According to this proposal, (3) constitutes a discursive disagreement because Bert's utterance expresses some conative attitude that disagrees with the conative attitude expressed by Ava's utterance.

---

<sup>24</sup>See the references in fn.7.

<sup>25</sup>This option is open to both contextualists and relativists, and has been advocated by both. See Egan 2012, 2014; Sundell 2011; Huvenes 2012; Marques and García-Carpintero 2014.

Both ways of developing this strategy face two serious worries. The first concerns the scope of the strategy. Even if some subjective disagreements involve disagreement in attitude, it's far from clear that all of them do. As we've seen, there seem to be subjective disagreements (both discursive and speechless) involving epistemic modals—for example, the lieutenant's disagreement with the detective over whether the butler might be guilty. But it's dubious that there is any clash in conative attitudes here. Of course, we can imagine a version of the case where there is such a disagreement in attitude—for example, we can imagine that the lieutenant wants the detective to share her belief that the butler might be guilty, whereas the detective has a conflicting desire. But it's by no means obvious that such disagreements involving epistemic modals *necessarily* involve such conflicting desires. The worry, then, is that even if appealing to disagreement in attitude helps explain some cases of subjective disagreement—e.g., those involving gustatory or moral matters—it won't generalize to handle others.<sup>26</sup>

A second worry arises when we ask, 'What does it take for two attitudes to disagree?' One initially tempting option is to explain disagreement in attitude in terms of whether the contents of the attitudes are consistent:

**Content Inconsistency** Two conative attitudes disagree iff they have inconsistent contents.

This is just the Simple Account of doxastic disagreement extended to conative attitudes. And it faces much the same difficulty: it overgenerates disagreements. To see this, suppose that after dinner Ava and Bert share a taxi. Bert wants to sit in front; Ava wants to sit in back. According to the standard treatment, the content of Bert's desire is the centered proposition:  $\{\langle w, i \rangle \mid i \text{ sits in the front at } w\}$ ; the content of Ava's desire is an inconsistent centered proposition:  $\{\langle w, i \rangle \mid i \text{ sits in the back at } w\}$ .<sup>27</sup> Given Content Inconsistency, their desires disagree. But this is clearly wrong: intuitively, there's no disagreement here.<sup>28</sup>

---

<sup>26</sup>And it may just turn out that whatever is the right story about other cases of doxastic disagreement will also extend to cases involving taste predicates and moral vocabulary, rendering appeals to disagreement in attitude otiose.

<sup>27</sup>The main motivation for this approach is the observation that desires seem to be *de se*. A familiar example from [Morgan 1970](#) makes this point: Joe is running for office, but secretly yearns to lose. In a drunken stupor, he is watching a political candidate on television; impressed with the candidate's oratory, he thinks: *I hope that guy wins*. As a matter of fact, the candidate in question is none other than Joe. Intuitively, the following is false in this situation:

(ii) Joe wants to win.

Those attracted to the Lewisian approach to the *de se* have thus been lead to hold that the contents of desires are centered propositions. (For relevant discussion, see [Chierchia 1989](#); [Ninan 2010](#).)

<sup>28</sup>Cf. [MacFarlane \(2014\)](#): chp.6's discussion of "non-cotenability". Oddly, despite raising a similar example, MacFarlane describes non-cotenability as sufficient for a type of disagreement. My intuition, however, is that when it comes to Bert and Ava's taxi-seating preferences, there's no sense in which they disagree. (We can, of course, stipulatively define such a sense of disagreement, but it doesn't seem that they disagree in any natural, pre-theoretic sense of the term.)

Faced with this difficulty, some may observe that Ava and Bert's seating preferences are jointly satisfiable: if Bert sits in the front and Ava sits in the back, both will get what they want. Perhaps, then, disagreement in attitude amounts to having conative attitudes that cannot be jointly satisfied:

**No Joint Satisfiability** Suppose A has a conative attitude with content  $p$  and B has a conative attitude with content  $q$ . These attitudes disagree iff they cannot be jointly satisfied—i.e.,  $\nexists w$  s.t.  $\langle w, A \rangle \in p$  and  $\langle w, B \rangle \in q$ .<sup>29</sup>

This is just the conative counterpart of No Joint Accuracy. And just as No Joint Accuracy undergenerates disagreements, so does No Joint Satisfiability. Consider again Ava and Bert's dining disagreement. Those who wish to diagnose this as a disagreement in attitude will presumably explain this in terms of a conflict in Ava and Bert's desires. What are the relevant desires? Presumably, the idea is that Ava desires to continue savoring the taste of the escargot, whereas Bert desires to be rid of its taste (or something along these lines). But these desires are jointly satisfiable: all that's required is for Ava to continue eating and for Bert to refrain. And so No Joint Satisfiability predicts that there's no disagreement in attitude.<sup>30</sup>

Therefore, those who seek to explain subjective disagreement in terms of disagreement in attitude face a version of the dilemma that arose in §2: they can either explain disagreement in attitude in terms of inconsistency in contents, in which case they overgenerate disagreements, or they can explain disagreement in attitude in terms of the preclusion of joint satisfiability, in which case they undergenerate disagreements. Is there any way out of this dilemma? One option—suggested by Marques and García-Carpintero—is to retain No Joint Satisfiability but rethink the contents of the desires (2014: 718). Perhaps Ava not only desires to continue enjoying the taste of the escargot, but also that Bert comes to appreciate the pleasures of gastropod gastronomy. More generally, Marques and García-Carpintero suggest that we often have '*de nobis*' desires—desires and to collectively embark on some common course of action, or to converge in our tastes. In scenarios such as the dining dispute, these *de nobis* desires will not be jointly satisfiable.

---

<sup>29</sup>This has good claim to being the standard account of disagreement in attitude. It offers a natural way of understanding Stevenson's restaurant example, and it has been defended—with varying degrees of explicitness—in Hall 1947; Stevenson 1963; Jackson 1985; Lewis 1989; Blackburn 1998; Marques and García-Carpintero 2014.

<sup>30</sup>This example suggests that No Joint Satisfiability does not provide sufficient conditions for disagreement in attitude (at least, if disagreement in attitude is supposed to explain speechless disagreement). One might also worry that No Joint Satisfiability fails to provide necessary conditions for disagreement in attitude. Ridge (2013) offers the example of two prospective renters, both of whom wish to rent the same flat. Their desires are not jointly satisfiable, but—according to Ridge—it's not obvious that we should report them as disagreeing, since both agree about all of the flat's merits.

It is certainly true that we often have such *de nobis* desires. But need they always be present in cases of speechless disagreement? Imagine a version of the dining dispute in which Ava doesn't care whether Bert comes to appreciate the taste of the molluscs: upon noticing his displeasure, she shrugs and thinks, 'To each his own.' Despite the absence of any jointly unsatisfiable *de nobis* desires, it seems they still disagree about whether the escargot is tasty. Of course, this disagreement is not particularly serious or important. And it may well be that every *serious* case of disagreement in attitude—every case that is likely to generate inter-personal conflict or hamper attempts to coordinate our actions—is a case where jointly unsatisfiable *de nobis* desires are present. But just because a disagreement isn't serious doesn't mean it isn't a disagreement.

Another option is to rethink the conative attitudes themselves. Thus far we've assumed that the conative attitudes involved in disputes about taste are desires, which have centered propositions as their objects. But this isn't the only possibility. [Baker and Woods \(2015\)](#) observe that the attitudes of *liking the taste of escargot* and *disliking the taste of escargot* seem to disagree with one another, even though the objects of these attitudes would appear to be an entity (the taste of escargot), not a proposition (centered or otherwise).<sup>31</sup> Perhaps, then, we should think of Ava and Bert's disagreement in attitude along these lines.

In response, we should first note that even if Baker and Woods are right that *liking x* and *disliking x* are attitudes that disagree, we still lack an account of this disagreement. What explains why these two attitudes disagree, if not the inconsistency or joint unsatisfiability of their contents? This is by no means obvious. Until some story is given, we won't have accomplished our goal of providing a fully general theory of subjective disagreement.

Second, it seems implausible that every speechless disagreement involving matters of taste can be handled in this way. Consider yet another variant of our dining dispute: suppose that Bert actually enjoys the taste of escargot, but is under the delusion that he loathes it. Prior to trying a bite, he disagrees with Ava about whether it is tasty: she believes it is; he believes it isn't. But we cannot diagnose their disagreement as one that arises from conflicting states of *liking the taste of escargot* and *disliking the taste of escargot* since both enjoy its taste.

In summary, the strategy of explaining subjective disagreement as disagreement in attitude faces serious difficulties. First, even if this strategy succeeds in explaining our intuitions about some subjective disagreements, it seems unlikely to generalize to all subjective disagreements (e.g., those involving epistemic modals). Second, in order for this strategy to succeed, it needs to be supplemented with an account of disagreement in

---

<sup>31</sup>Of course, it is debatable whether appearances are to be trusted here; at least some have argued that *liking* and *disliking* should be analyzed as relations to propositions after all. For relevant discussion, see [Sinhababu 2015](#); [Grzankowski 2012, 2016](#).

attitude. But when we try to provide such an account, we arrive at a familiar impasse. If we try to understand disagreement in attitude in terms of inconsistency in contents, we get too much disagreement; if we try to understand it in terms of joint unsatisfiability, we get too little.

## 5 A New Approach

Let's take stock. We started with the natural idea that disagreement is a matter of conflicting beliefs: objective disagreements occur when one party believes a possible worlds proposition and another disbelieves it; subjective disagreements occur when one party believes a centered proposition and another disbelieves it. The problem with this proposal was that it overgenerated disagreements among ordinary *de se* beliefs. This led us to explore other strategies for understanding subjective disagreement: treating it as a discursive phenomenon, or explaining it in terms of disagreement in attitude. Both of these strategies ran into difficulties. And it seems that at least part of the difficulty arose from abandoning our initial idea that disagreement is at heart a doxastic phenomenon. This explains why the discursive strategy had trouble accounting for speechless disagreements, as well as why the disagreement in attitude strategy had trouble accounting for the full range of subjective disagreements (e.g., those involving epistemic modals).

In this section, I lay out a new strategy for understanding disagreement. The strategy retains our starting idea that all disagreement—both objective and subjective—is fundamentally a doxastic phenomenon. However, it denies that all doxastic disagreement is a function of doxastic *contents*. According to the approach developed here, our intuitions about whether two beliefs disagree are sensitive to syntactic features of the sentences used to report them.

### 5.1 An Analysis of Disagreement Ascriptions

Thus far, I've been proceeding in the material mode: I've appealed to intuitions about the circumstances under which agents agree or disagree, and used these intuitions as data guiding my pursuit of a theory of disagreement. In developing my account, it will be useful to transition to the formal mode: rather than appealing to intuitions about disagreement, let us appeal to intuitions about *disagreement ascriptions*—that is, sentences reporting disagreement. While there are different ways of ascribing disagreement, I will take the following schema as my paradigm:

- (4) A disagrees with B over whether  $\phi$ .

(Here  $\phi$  is a variable ranging over declarative sentences.)

We find it natural to use this schema to report instances of both objective and subjective disagreement:

- (5) Ava disagrees with Bert over whether global warming has human causes.
- (6) Ava disagrees with Bert over whether the escargot is tasty.
- (7) The lieutenant disagrees with the detective over whether the butler might be guilty.

For now, rather than providing an account of what disagreement is, let us instead pursue the related task of providing truth conditions for such disagreement ascriptions.<sup>32</sup>

What would such truth conditions look like? I wish to defend a very simple proposal:

**Disagreement Ascriptions:** ‘A disagrees with B over whether  $\phi$ ’ is true in a context  $c$  iff  
‘A believes  $\phi$ ’ and ‘B believes  $\neg\phi$ ’ are both true in  $c$  (or *vice versa*).

This analysis closely resembles the Simple Account of Disagreement. Like the Simple Account, it makes disagreement a fundamentally doxastic phenomenon. As a result, Disagreement Ascriptions—like the Simple Account—validates the following inferences:

- (8) a. Ava believes that global warming has human causes.  
b. Bert believes that global warming doesn’t have human causes.  $\Rightarrow$   
c. Ava disagrees with Bert over whether global warming has human causes.
- (9) a. Ava believes that escargot is tasty.  
b. Bert believes that escargot isn’t tasty.  $\Rightarrow$   
c. Ava disagrees with Bert over whether escargot is tasty.
- (10) a. The lieutenant believes the butler might be guilty.  
b. The detective believes that the butler can’t be guilty.  $\Rightarrow$   
c. The lieutenant disagrees with the detective over whether the butler might be guilty.

Since these inferences seem highly compelling, this is an important mark in favor of Disagreement Ascriptions.<sup>33</sup>

---

<sup>32</sup>Of course, it is often harmless to transition from the formal to the material mode. But not always—consider, for example, a semantic analysis of a context-sensitive expression. For reasons that will soon be clear, I think that disagreement ascriptions are among the problem cases: once we have our truth conditions, we cannot simply semantically descend, framing everything in the material mode.

<sup>33</sup>Note that none of the other rivals to the Simple Account validate all of these inferences. In particular, (9) and (10) seem to be valid regardless of whether the parties ever converse—a fact that is problematic for the Discursive Account. And, as we’ve seen, the Disagreement in Attitude Strategy does not generalize cleanly to handle disagreements involving epistemic modals; thus it does not validate (10).

At the same time, Disagreement Ascriptions avoids our objection to the Simple Account. Let us briefly recap our objection (§2): the Simple Account, when combined with the standard Lewisian analysis of the *de se*, overpredicts disagreements among ordinary *de se* beliefs. To refer back to our example: Kim believes HUNGRY (i.e., the set of centered worlds such that the center is hungry); Sara, who has just eaten, believes  $\neg$ HUNGRY. As we saw, the Simple View wrongly predicts that they disagree.

To see how Disagreement Ascriptions avoids this difficulty, consider: how would we report Kim's belief in HUNGRY, and Sara's belief in its negation? The natural way to do so would be as follows:

- (11) Kim believes (/thinks) she is hungry.
- (12) Sara believes (/thinks) she isn't hungry.

But a moment's reflection reveals that these belief ascriptions are ambiguous, and only attribute *de se* beliefs on one of their readings. For example, suppose that the occurrence of *she* in (11) is taken to refer to Sara rather than Kim. Then (11) would no longer attribute a *de se* belief; rather, it would attribute a *de re* belief about Sara. (The same holds, *mutatis mutandis*, if the occurrence of *she* in (12) is taken to refer to Kim.)

This suggests that the *de se* interpretations of (11) and (12) constrain the resolution of the pronoun in the complement clause. Such constraints have long been recognized in the semantics literature on *de se* attitude ascriptions. While there are various semantic treatments of such ascriptions, the leading approach is due to Chierchia (1989). According to Chierchia, *de se* readings of attitude ascriptions are associated with distinctive logical forms. In particular, the *de se* readings of an attitude ascription is associated with an LF that contains a phonologically null abstraction operator O, which is co-indexed with both the pronoun and the attitude holder.<sup>34</sup> Thus the *de se* readings of (11) and (12) are associated with the following LFs:

- (11')  $\text{Kim}_i$  believes (/thinks)  $[O_i \text{ she}_i \text{ is hungry}]$ .
- (12')  $\text{Sara}_k$  believes (/thinks)  $[O_k \text{ she}_k \text{ isn't hungry}]$ .

This approach entails the following constraint on the resolution of pronominal elements in the complement clauses of *de se* attitude reports:

**Co-Indexing Requirement** In order for a pronoun in the complement clause of an attitude report to attribute a *de se* belief, the pronoun must be co-indexed with the attitude

---

<sup>34</sup>I discuss the semantics for this operator in §6.

holder.<sup>35</sup>

As Chierchia observes, the Co-Indexing Requirement has independent explanatory appeal. In particular, it explains why a pronoun in a *de se* attitude report must agree with the attitude holder in gender and number (1989: 10). Consider, for example:

- (13) Kim believes that he is hungry.

If Kim is a woman, (13) does not have an available *de se* reading.<sup>36</sup> Similarly, the sentence:

- (14) Kim believes that they are hungry.

also lacks a *de se* reading. The Co-Indexing Requirement explains these observations.

Given the Co-Indexing Requirement, Disagreement Ascriptions correctly predicts that we cannot truthfully ascribe any disagreement to Kim and Sara. Consider: how might we ascribe such a disagreement? Presumably, through an utterance such as the following:

- (15) Kim disagrees with Sara over whether she is hungry.

But regardless of how we resolve the pronoun *she* in the complement clause, we cannot infer (15) from (11') and (12'). To do so would be to commit a fallacy of equivocation, akin to the inference in (16):

- (16) a. John believes it's raining here. (*Uttered in NYC*)  
b. Fred believes it isn't raining here. (*Uttered in LA*)  $\neq$   
c. John disagrees with Fred over whether it's raining here.

To take stock: my analysis of disagreement ascriptions is able to capture our intuitions of disagreement in a wide variety of cases—both objective and subjective. At the same time, it avoids overpredicting disagreement among ordinary *de se* beliefs. It is worth reflecting on how my analysis manages to achieve this result. Whereas the Simple Account analyzed disagreement in terms of the inconsistency of the contents believed, my account analyzes disagreement ascriptions in terms of the inconsistency of the *sentences* used to characterize beliefs. In doing so, my account is able to appeal to certain syntactic features of ordinary

---

<sup>35</sup>This co-indexing is a necessary condition, not a sufficient condition. On Chierchia's approach, there is also a distinct *de re* reading of (11) given by the LF:

(iii)  $\text{Kim}_i$  believes (/thinks) [ $\text{she}_i$  is hungry].

Here the pronoun is co-indexed with the attitude holder, but the ascription is not interpreted *de se*, since it lacks the abstraction operator O.

<sup>36</sup>See Schlenker 2003: 78-79 for relevant discussion.

*de se* belief ascriptions—in particular, the Co-Indexing Requirement—to explain why such ascriptions do not entail a corresponding disagreement ascription.

What are the contents of *de se* beliefs, according to my analysis? As formulated, my analysis is officially neutral on how to analyze *de se* beliefs. However—for reasons that will become apparent in §6—it is perfectly compatible with the standard Lewisian treatment of the *de se*. It is thus compatible with saying that the content of Kim’s *de se* belief is HUNGRY, and the content of Sara’s *de se* belief is  $\neg$ HUNGRY. And if we do say this, then we arrive at a surprising result: the consistency relations that obtain between the contents of agents’ beliefs can come apart from the consistency relations that obtain between the sentences used to characterize those beliefs.

If this result is correct, we should expect to find cases where describing the same belief using different sentences affects our intuitions about disagreement. I’ll now argue that this is precisely what we find.

## 5.2 Further Data

Compare (9a) (repeated here as (17)) with (18):

- (17) Ava believes the escargot is tasty.
- (18)  $Ava_i$  believes the escargot is tasty to her<sub>i</sub>.

Both of these belief reports have a natural *de se* reading.<sup>37</sup> Given the Lewisian view of the *de se*, on this reading (17) and (18) are equivalent: both say that Ava believes TASTY (that is, the set of centered worlds such that the center enjoys the taste of the escargot).

While these two belief reports are naturally read as equivalent, they generate strikingly different intuitions about disagreement. Consider again snail-loathing Bert. It’s natural to report him as disagreeing with the belief reported by (17):

- (19) Bert disagrees with Ava over whether the escargot is tasty.

It’s far less natural to report him as disagreeing with the belief reported in (18):

- (20) Bert disagrees with  $Ava_i$  over whether the escargot is tasty to her<sub>i</sub>.

More generally, (19) and (20) seem to have very different truth conditions. Intuitively, (19) is true as long as Ava believes the escargot is tasty and Bert believes it isn’t. By contrast,

---

<sup>37</sup>We’ve already seen the motivation for this in §2: consider the scenario in which Ava knows that she hates the taste of the escargot, but mistakenly believes *that woman over there* enjoys its taste, where *that woman over there* is none other than Ava. There seem to be readily available readings of both (17) and (18) on which they are false in this scenario.

(20) would only be true if Bert thinks that Ava is mistaken about whether she enjoys the taste of the escargot.

This difference in the truth conditions of (19) and (20) is surprising, and is not predicted by any of the theories of disagreement canvassed in §§2-4. For example, *De Se* Conflict wrongly predicts that (19) and (20) are both true. (After all, the content of Ava's belief that the escargot is tasty to her is just TASTY; according to *De Se* Conflict, Bert disagrees with this content.) The Discursive Account also fails to explain the difference in the truth conditions of (19) and (20), since this difference seems to be unrelated to whether Ava and Bert are engaged in a discursive disagreement. (In our speechless dining disagreement scenario, it still seems that (19) is true and (20) is false.) And the Disagreement in Attitude Strategy fares no better: if all it is for Ava to believe the escargot is tasty is for her to believe that the escargot is tasty to her, then whatever disagreement in attitude arises in virtue of Ava believing the escargot is tasty should also arise in virtue of her believing that it's tasty to her.

By contrast, my analysis of disagreement ascriptions straightforwardly explains this data. According to my analysis, (19) is true since the relevant belief ascriptions ((17) (*Ava believes the escargot is tasty*) and (9b) (*Bert believes the escargot isn't tasty*)) are both true. However, in order for (20) to be true, the following would also have to be true:

- (21) Bert believes the escargot isn't tasty to her. (where *her* refers Ava)

Thus my analysis correctly predicts the difference in the truth conditions of (19) and (20). Once again, the reason my analysis succeeds where others fail is that my analysis allows for certain syntactic features of belief ascriptions (specifically, the presence of a pronoun that is co-indexed with the believer) to affect the truth or falsity of disagreement ascriptions.

A similar pattern can be found in attitude reports involving epistemic modals. Compare:

- (22) The lieutenant believes the butler might be guilty.  
(23) The lieutenant<sub>i</sub> believes that, for all she<sub>i</sub> knows, the butler might be guilty.

As we've seen (§2), both of these reports have readily accessible *de se* readings. On this reading, (22) and (23) are equivalent: both say that the lieutenant believes GUILTY (that is, the set of centered worlds such that the butler's guilt is compatible with what the center knows). Despite this, these belief reports generate very different intuitions about disagreement. In particular, it's far more natural to report the detective (who has conclusively ruled out the butler as a suspect) as disagreeing with the belief reported

in (22) than it is to report the detective as disagreeing with the belief reported in (23). Compare:

- (24) The detective disagrees with the lieutenant over whether the butler might be guilty.
- (25) The detective<sub>k</sub> disagrees with the lieutenant<sub>i</sub> over whether, for all she<sub>i</sub> knows, the butler might be guilty.

(24) seems true. By contrast, (25) would only be true in a situation where detective has views on whether the butler's guilt is compatible with the lieutenant's knowledge.

Once again, we have a case where the truth or falsity of a disagreement ascription is not determined by the contents of the beliefs in question. And once again, my analysis explains the data. (24) is true because the corresponding belief ascriptions are both true. (25) is false, because (26) is false when *she* is taken to refer to the lieutenant:

- (26) The detective believes that, for all she knows, the butler can't be guilty.

### 5.3 Taking Stock

In this section, I've offered a new way of approaching disagreement. The guiding idea is to reframe our task by way of semantic ascent, and focus on providing an analysis of disagreement ascriptions. I went on to offer a simple analysis of disagreement in terms of inconsistent belief ascriptions, where the inconsistency is located not in the contents of the beliefs, but rather in the sentences used to characterize them. I argued that this approach avoids the difficulties facing the other accounts of disagreement we've considered. In particular, it gives us enough subjective disagreement, without giving us too much. Moreover, I argued that one surprising consequence of this approach—that the truth conditions for disagreement ascriptions are not determined entirely by the contents of the parties' beliefs—is supported by independent data (§5.2), data that creates independent difficulties for the other accounts of disagreement we've considered. Given this, it seems there's reason to hope that my analysis—or something very much like it—is on the right track.

However, thus far I've left many details undeveloped. In particular, while I've helped myself to intuitions about the truth conditions for various belief ascriptions, I have not offered a semantics for belief ascriptions that predicts these intuitions. More importantly, I have not shown how my analysis of disagreement can be implemented compositionally. This lacuna may appear particularly worrisome: if the truth conditions for disagreement ascriptions are not determined by the contents of the beliefs in question, it might seem that our lexical entry for *disagree* will have to take as arguments *belief reports*, thereby violating compositionality. Finally, I have not said anything about how my analysis of

disagreement ascriptions bears on the semantics of subjective expressions: does it count in favor of relativism, or contextualism, or neither? In the next section, I offer one—but by no means the only—way of fleshing out the details. (Readers uninterested in the formal nuts-and-bolts should feel free to skip ahead.)

## 6 Filling in the Details

In this section, I develop a view according to which the *de se* readings of attitude ascriptions involving overt pronouns (e.g., (11), (12), (18), (23)) function differently than the *de se* readings of attitude ascriptions that lack overt pronouns (e.g., (17), (22)). To explain the former, I make use of Chierchia's (1989) approach, which posits a null abstraction operator subject to the Co-Indexing Requirement. To explain the latter, I make use of Stephenson's (2007a) approach, which appeals to a special judge-sensitive pronoun ( $\text{PRO}_J$ ) as well as a special composition rule for handling judge-sensitive expressions. According to the view that results, attitude ascriptions involving overt pronouns are subject to syntactic restrictions that do not apply to attitude ascriptions that lack overt pronouns. I show that this difference in treatment gives us the right results: by combining this with a simple compositional semantics for *disagree*, we can explain why the *de se* readings of attitude ascriptions lacking overt pronouns generate intuitions of disagreement, whereas the *de se* readings of attitude ascriptions containing overt pronouns do not. Finally, I show that while Stephenson (2007a) employed a relativist framework, this feature is inessential: we can devise a contextualist counterpart to Stephenson's (2007a) semantics that delivers the exact same predictions.

### 6.1 Attitude Ascriptions Containing Overt Pronouns

While Lewis (1979) offered an account of the contents of *de se* attitudes, he never offered a compositional semantics for attitude reports. Chierchia (1989) provides what has good claim to be the canonical formal implementation. As we've seen, Chierchia takes the *de se* reading of (11) (*Kim believes she is hungry*) to be given by the LF in (11'):

- (11') Kim<sub>i</sub> believes [<sub>CP</sub> O<sub>i</sub>. she<sub>i</sub> is hungry].

As before, O is an abstraction operator co-indexed with both the attitude holder and the pronoun in the attitude report (*she*). How should we analyze this operator? First, let us adopt a standard analysis of pronouns as variables whose values are supplied by a contextually determined variable assignment function  $g$  (I omit gender presuppositions for the sake of simplicity):

$$[\![\text{she}_i]\!]^{c,g,w} = g(i).$$

We can then formulate the semantics for  $O$  as follows:

$$[\![O_i \alpha]\!]^{c,g,w} = \lambda x_e. [\![\alpha]\!]^{c,g[x/i],w}$$

Here  $g[x/i]$  is a modified variable assignment function: it's a variable assignment function that is exactly like  $g$ , except that it maps anything with index  $i$  to  $x$ .

There remains the question of the semantics for *believes*. Here I adopt a standard extension of a Hintikka semantics for *believes*, according to which *believes* is a universal quantifier over an agent's doxastic alternatives (Hintikka 1962). Since we wish to implement a Lewisian view of the *de se*, we will model doxastic alternatives as sets of centered worlds:

$$[\![\text{believes}]\!]^{c,g,w} = \lambda p_{\langle s,et \rangle}. \lambda x_e. \forall \langle w', x' \rangle \in \text{Dox}(w, x) : p(w')(x') = 1,$$

where  $\text{Dox}(w, x) = \{ \langle w', x' \rangle \mid \text{being } x' \text{ at } w' \text{ is compatible with what } x \text{ believes at } w \}$ .

By combining these ingredients, we are able to derive what—from the Lewisian point of view—are the correct truth conditions for (11'). In particular, we are able to derive the result that (11') is true iff Kim believes HUNGRY. (Throughout this section, the derivations of the truth conditions will be relegated to appendix A.)

Similarly, on this approach the *de se* reading of (12) (*Sara believes she isn't hungry*) will be associated with the LF provided in (12'):

$$(12') \quad \text{Sara}_k \text{ believes } [\![CP \ O_k \ \text{she}_k \ \text{isn't hungry}]\!].$$

By parallel reasoning, we get the result that (12') is true iff Sara believes  $\neg$ HUNGRY.

## 6.2 A Compositional Semantics for *Disagree*

How should we formulate our lexical entry for *disagree*? I propose a simple answer: *disagree* is a function that takes as arguments two agents and a centered proposition  $p$  (boring or otherwise), and yields the value 'true' if and only if one of the agents stands in the 'believes' relation to  $p$  and the other agent stands in the 'believes' relation to  $\neg p$ . More precisely:

$$[\![\text{disagree}]\!]^{c,g,w} = \lambda p_{\langle s,et \rangle}. \lambda x_e. \lambda y_e. [\![\text{believe}]\!]^{c,g,w}(p)(x) = [\![\text{believe}]\!]^{c,g,w}(\lambda w'_s. \lambda z_e. p(w')(z) = 0)(y) = 1, \text{ or } [\![\text{believe}]\!]^{c,g,w}(\lambda w'_s. \lambda z_e. p(w')(z) = 0)(x) = [\![\text{believe}]\!]^{c,g,w}(p)(y) = 1.$$

This lexical entry is fully compositional: *disagree* does not take the LFs of belief reports as arguments; rather, its arguments are individuals and centered propositions. One might then wonder: how does this tell my view apart from the Simple Account of Disagreement?

The answer is that nothing in the lexical entry for *disagree* distinguishes my view from the Simple Account. The difference comes when we try to ascribe disagreements to two individuals. After all, it takes more than a word to ascribe disagreement: a full ascription will take the form of a declarative sentence (e.g., an instance of (4)). And when we try to ascribe disagreements to two individuals with incompatible *de se* beliefs, we find ourselves hampered by the syntactic constraints on the disagreement reports.

We are now in a position to see this more clearly. Consider again our failed attempt to ascribe disagreement to Kim and Sara ((15), reprinted here as (27)):

- (27) Kim disagrees with Sara over whether she is hungry.

What is the LF of (27)? Since (27) purports to report a disagreement in *de se* beliefs, we need to associate the pronoun in the complement clause (*she*) with our null operator O, which is co-indexed with the believer. But since we have two believers, there are two possible LFs:

- (27) a. Kim<sub>i</sub> disagrees with Sara<sub>k</sub> over whether [<sub>CP</sub> O<sub>i</sub> she<sub>i</sub> is hungry].  
b. Kim<sub>i</sub> disagrees with Sara<sub>k</sub> over whether [<sub>CP</sub> O<sub>k</sub> she<sub>k</sub> is hungry].

But of course the pronoun in the complement clause of (11') is associated with a different index than the pronoun in the complement clause of (12'). Thus our lexical entry for *disagree* ensures that we cannot infer either reading of (27) from (11') and (12'). Thus we have made good on our initial diagnosis that any attempt do so involves a fallacy of equivocation.

### 6.3 Attitude Ascriptions Without Pronouns

We have seen one way of deriving truth conditions for *de se* attitude reports: positing a null abstraction operator. Note that this strategy does not yet say anything about subjective expressions: it does not give us truth conditions for attitude reports involving taste predicates, epistemic modals, or the like. In this subsection, I review one strategy for providing such truth conditions, due to Stephenson (2007a).

Stephenson's semantics is relativist: she supplements the index with a separate judge parameter *j*. (I'll shortly be arguing that this is a dispensable feature of Stephenson's approach.) Stephenson offers the following lexical entry for *tasty*:

$$\llbracket \text{tasty} \rrbracket^{c,g,w,j} = \lambda x_e. \lambda y_e. y \text{ tastes good to } x \text{ at } w.$$

While this lexical entry includes a judge parameter, it is not yet relativist, since the semantic value of *tasty* does not depend directly on the judge parameter. The relativist

dimension comes in when Stephenson posits a null pronominal entity  $\text{PRO}_J$  that refers to the judge:

$$\llbracket \text{PRO}_J \rrbracket^{c,g,w,j} = j.$$

Combining these two lexical entries we get:

$$\llbracket \text{tasty PRO}_J \rrbracket^{c,g,w,j} = \lambda y_e. y \text{ tastes good to } j \text{ at } w.$$

According to Stephenson, while *tasty* can take  $\text{PRO}_J$  as an argument, the value of  $x$  can also be supplied by prepositional phrases, e.g.:

- (28) The escargot is tasty to Ava.<sup>38,39</sup>

Let us see how this approach can be combined with Chierchia's approach (§6.1), together with our semantics for *disagree* (§6.2) in a way that explains our data from §5.2. Recall our pair of belief ascriptions:

- (29) Ava believes the escargot is tasty.  
 (30) Ava believes the escargot is tasty to her.

Following Stephenson, let us assume that the *de se* reading of (29) is associated with the following LF:

- (31) Ava believes [<sub>CP</sub> the escargot is tasty  $\text{PRO}_J$ ].

To compute the truth conditions for (31), we need one final tool: a modified version of intensional function application, tailored to handle expressions with judge-sensitive semantic values. Stephenson offers the following:

**IFA\*** If  $\alpha$  is a complex expression formed by combining two expressions  $\beta$  and  $\gamma$ , and  
 $\lambda w'_s. \lambda j'_e. \llbracket \gamma \rrbracket^{c,g,w',j'}$  is in the domain of  $\llbracket \beta \rrbracket^{c,g,w,j}$ , then  $\llbracket \alpha \rrbracket^{c,g,w,j} = \llbracket \beta \rrbracket^{c,g,w,j} (\lambda w'_s. \lambda j'_e. \llbracket \gamma \rrbracket^{c,g,w',j'})$ .

As Stephenson shows, by combining the relativist semantics for *tasty*  $\text{PRO}_J$  with IFA\* and our semantics for *believes* (§6.1), we derive the result that taste belief is *de se*. More precisely, we derive the result that (32) is true iff Ava believes TASTY.

---

<sup>38</sup>Stephenson also holds that the value of  $x$  can be supplied by contextually salient individuals. (This is to allow for exocentric uses of taste predicates.) Pearson (2013) convincingly argues that the mere fact that an individual is contextually salient does not suffice for that individual to become available as an argument for *tasty*. Perhaps a better proposal would be to say that the value of  $x$  can be supplied by other contextually salient individuals, provided that the speaker has manifested an intention to regard those individuals' tastes as relevant to the interpretation of the taste predicate.

<sup>39</sup>For complementary arguments that *tasty* takes an experiencer argument, see Glanzberg 2007 and Schaffer 2011. For reservations, see Richard 2008: chp.5.

Now, what LF should we assign (30)? In (30), the argument for *tasty* is supplied by a prepositional phrase containing an overt pronoun. Thus to analyze (30), we wheel in the resources of Chierchia's approach. In particular, we take its *de se* reading to be provided by an LF containing our covert abstraction operator O:

- (32)  $\text{Ava}_i \text{ believes } [CP \text{ O}_i \text{ the escargot is tasty to her}_i]$ .

By combining Stephenson's semantics for *tasty* with our semantics for O and our semantics for *believes*, we derive the result that (32) is also true iff Ava believes TASTY. (See Appendix A.2 for the details.) And so the present approach predicts that (29) and (30) both have equivalent *de se* readings, as desired.

At the same time, we also predict that (29) and (30) license different disagreement reports. Specifically, we predict that (33) is true but (34) is false in our dining dispute:

- (33) Bert disagrees with Ava about whether the escargot is tasty.  
 (34)  $\text{Bert}_i \text{ disagrees with } \text{Ava}_k \text{ about whether the escargot is tasty to her}_k$ .

Start with (33). Since there are no prepositional phrases supplying an argument to *tasty*, we take its argument to be supplied by PRO<sub>J</sub>, yielding the following LF:

- (35) Bert disagrees with Ava about whether  $[CP \text{ the escargot is tasty PRO}_J]$ .

Now, we've seen that (31) (*Ava believes the escargot is tasty PRO<sub>J</sub>*) is true in our dining dispute. By parallel reasoning, *Bert believes the escargot is not tasty PRO<sub>J</sub>* is also true. And so, by our semantics for *disagrees*, (35) is true.

Next, consider (34). What LF should we assign this sentence? This is not obvious; given the co-indexing requirement, there is no way to construe the pronoun in the complement clause as ascribing a *de se* belief to Bert. A natural option is to interpret the sentence as ascribing a *de re* disagreement:

- (36)  $\text{Bert}_i \text{ disagrees with } \text{Ava}_k \text{ about whether } [\lambda x. \text{ the escargot is tasty to her}_k]$ .

But (36) is false, and our theory predicts as much. Of course, the relevant *de re* belief ascription to Ava:

- (37)  $\text{Ava}_k \text{ believes } [\lambda x. \text{ the escargot is tasty to her}_k]$ .

is true (assuming that *de se* beliefs entail their *de re* counterparts). However, the relevant *de re* belief ascription to Bert:

- (38)  $\text{Bert}_i \text{ believes } [\lambda x. \text{ the escargot is not tasty to her}_k]$ .

is false when evaluated relative to a variable assignment function that maps  $k$  to Ava.

Where does this leave us? Thus far this section has discussed two treatments of *de se* attitude ascriptions. The first—due to Chierchia—covers *de se* attitude ascriptions involving overt pronouns. This treatment involved a syntactic restriction (the Co-Indexing Requirement) governing the relation between the subject of the ascription, the overt pronoun, and a covert abstraction operator. The second treatment—due to Stephenson—covered *de se* attitude ascriptions lacking overt pronouns. This treatment was free from any syntactic restrictions; it relied instead on lexical entries that optionally accept a judge argument, as well as a special composition rule (IFA\*). By synthesizing these two treatments, we predict that there is an important difference between *de se* attitude ascriptions that involve overt pronouns and those that do not. The difference resides not in the contents of the attitude ascriptions, but rather their sensitivity to syntactic restrictions. By exploiting this difference, we get the result we sought: we predict enough disagreement (e.g., we predict the truth of (33)), without overgenerating disagreement (in particular, we avoid predicting the truth of either (27) or (36)).<sup>40</sup>

While I have focused on taste predicates in this section, the strategy extends to other subjective expressions. For example, the same strategy works for epistemic modals: following Stephenson (2007a), we can say that an epistemic modal optionally takes  $\text{PRO}_J$  as an argument; however, its argument can also be supplied by overt restrictor phrases (e.g., *for all she knows*). When an epistemic modal in an attitude report takes  $\text{PRO}_J$  as an argument, we use IFA\* to derive the *de se* truth conditions for the attitude report. When its argument is supplied by an overt restrictor phrase containing a pronoun, we instead use Chierchia's strategy to get *de se* truth conditions. While in each case the truth conditions will be equivalent, different disagreement reports will be licensed: in particular, we will predict true disagreement reports when the modal takes  $\text{PRO}_J$  as argument, and false disagreement reports when its argument is supplied by an overt pronoun. This

---

<sup>40</sup>Note that Stephenson does not herself appeal to Chierchia's strategy for handling ordinary *de se* beliefs. Indeed, Stephenson (2010) uses  $\text{PRO}_J$  to give an alternative explanation of the fact that control constructions are obligatorily *de se*. According to Stephenson's proposal, the LF of (ii) (*Joe wants to win*) is:

(ii') Joe wants [ $\text{PRO}_J$  to win].

To my knowledge, Stephenson does not discuss *de se* attitude ascriptions containing overt pronouns (e.g., (11)). However, if one were to try to extend her 2010 proposal to such ascriptions, we risk once again overgenerating disagreements. To see this, suppose we took the pronoun *she* in the complement clause of (11) to denote  $\text{PRO}_J$ , giving us the following LF:

(11'') Kim believes [ $\text{PRO}_J$  is hungry].

This proposal incorrectly predicts a true reading of (15), given by:

(iii) Kim and Sara disagree over whether [ $_{CP} \text{PRO}_J$  is hungry].

This suggests that we should not dispense with Chierchia's route to the *de se* in favor of Stephenson's. Rather, both routes are required to get disagreement ascriptions to come out right.

approach thus predicts that (22) and (23) are equivalent; however, (24) is true, whereas (25) is false. The strategy can be extended in similar fashion to other candidate subjective expressions—for example, moral and aesthetic expressions.

#### 6.4 A Contextualist Alternative

In the foregoing, we followed Stephenson in relying on a relativist framework. However, closer inspection reveals that the relativism isn't essential; Stephenson's basic strategy for compositionally implementing a *de se* semantics for subjective expressions in attitude reports can be transposed into a contextualist framework.

Let's walk through how this would work for taste predicates. Here is a minimal contextualist variant of Stephenson's lexical entry for *tasty*:

$$[\text{tasty}]^{c,g,w} = \lambda x_e. \lambda y_e. y \text{ tastes good to } x \text{ at } w.$$

In all important respects, this is the same as Stephenson's semantics; the only difference is that our semantic values are only relativized to variable assignments (supplied by the context of utterance), and worlds; there is no judge or assessor parameter.

Next, assume that the value of  $x$  is typically supplied by a special null pronominal entity  $\text{PRO}_\tau$ , whose value is supplied by the contextually determined variable assignment function, and which cannot be bound by the abstraction operator  $O$ :

$$[\text{PRO}_\tau]^{c,g,w} = g(\tau)$$

Combining these two lexical entries we get:

$$[\text{tasty PRO}_\tau]^{c,g,w} = \lambda y_e. y \text{ tastes good to } g(\tau) \text{ at } w.$$

While the value of  $x$  is typically supplied by  $\text{PRO}_\tau$ , as before we allow that its value can be supplied by overt PPs. Thus in (28) (*The escargot is tasty to Ava*) the value of  $x$  is Ava.

Can the contextualist use these ingredients to predict that taste belief is *de se*? Consider (29) (*Ava believes the escargot is tasty*). Assume that the *de se* reading of this sentence is given by the LF:

- (39) Ava believes [<sub>CP</sub> the escargot is tasty  $\text{PRO}_\tau$ ].

Computing the truth conditions for (39) using our lexical entry for *believes* and IFA delivers the wrong results. In particular, we get the result that (39) is true relative to  $g$  iff all of Ava's doxastic alternatives are such that the escargot tastes good to whatever individual is the value of  $g(\tau)$ . But for all we've said, the value of  $g(\tau)$  might very well be someone other than Ava, in which case (39) won't say that Ava believes TASTY. Indeed,

it is precisely this difficulty that motivates Stephenson and others to capture the *de se* dimension of taste beliefs in a relativist framework.

However, the contextualist has other options. Recall that Stephenson did not rely on IFA, but rather on a slightly different composition rule, IFA\*, tailored to handle judge-sensitive expressions. Thus it is also fair game for the contextualist to propose a modified composition rule. What the contextualist needs is the following: when we compute the truth conditions for (39) relative to some variable assignment function  $g$ , *believes* does not take as its argument the set of centered worlds where the complement clause is true as evaluated relative to  $g$ ; rather, it takes as its argument the set of centered worlds where the complement clause is true as evaluated relative to a modified variable assignment function  $g'$  that is exactly like  $g$  except that  $g'$  maps the center to  $\tau$ . That is:

**IFA\*\*** If  $\alpha$  is a complex expression formed by combining two expressions  $\beta$  and  $\gamma$ , and

$$\lambda w'_s. \lambda x_e. [\gamma]^{c,g,w'} \text{ is in the domain of } [\beta]^{c,g,w}, \text{ then } [\alpha]^{c,g,w} = [\beta]^{c,g,w}(\lambda w'_s. \lambda x_e. [\gamma]^{c,g[x/\tau],w'}).$$

A word of clarification. According to this composition rule,  $\text{PRO}_\tau$  has a privileged status among pronouns. In particular, the modified variable assignment function  $g[x/\tau]$  only binds  $\tau$ . This is important, since otherwise we would make incorrect predictions about *de re* belief ascriptions. For example, while (11) (*Kim believes she is hungry*) has a *de se* reading, it can also be read as ascribing a *de re* belief about someone else's hunger:

- (40) Kim<sub>i</sub> believes [λx. she<sub>k</sub> is hungry].

Suppose we evaluate (40) relative to a variable assignment function  $g$  that maps  $k$  to Sara. We do not want to predict that (40) is true, thus evaluated, iff Kim believes HUNGRY. But this is what we would be forced to say if we evaluated the complement clause relative to a modified variable assignment function that mapped anything with index  $k$  to  $x$  ( $g[x/k]$ ).<sup>41</sup>

Our clarification finished, we are in a position to see that IFA\*\* delivers what the contextualist needs. Even though the value of  $\text{PRO}_\tau$  is some individual supplied by the context of utterance (who may or may not be Ava), our lexical entry for *believes* joins forces with our special composition rule (IFA\*\*) in order to deliver the result (39) is true iff Ava believes TASTY. This is because combining our lexical entry for *believes* with IFA\*\* leads us to evaluate the embedded clause relative to a modified variable assignment function—one

---

<sup>41</sup>While the composition rule IFA\*\* is new, the basic idea of having a composition rule that evaluates a sentential constituent relative to a modified variable assignment function is familiar. For example, the standard analysis of relative clauses such as *who is hungry* relies on a rule of predicate abstraction, which evaluates moved structures using a modified variable assignment function (Heim and Kratzer 1998: 186). Thus IFA\*\* can be seen as integrating the ideas underlying traditional intensional function application and predicate abstraction in a single composition rule.

that maps the value of  $\tau$  to the centers of Ava's doxastic alternatives. (See Appendix A.3 for a derivation.)

What about (30) (*Ava believes the escargot is tasty to her*)? Here too the contextualist can follow the tracks laid down in §6.3. As before, we take the *de se* reading of this belief ascription to be provided by the LF in (32) (*Ava<sub>i</sub> believes O<sub>i</sub> the escargot is tasty to her<sub>i</sub>*). And, as before, this is in turn true if and only if Ava believes TASTY.

Like the relativist approach, the contextualist approach predicts that even though (39) is equivalent to (32), they license different disagreement reports. In particular, (33) (*Bert disagrees with Ava about whether the escargot is tasty*) will be associated with the LF:

- (41) Bert disagrees with Ava about whether [<sub>CP</sub> the escargot is tasty PRO <sub>$\tau$</sub> ].

which is predicted to be true. By contrast, (34) (*Bert<sub>i</sub> disagrees with Ava<sub>k</sub> about whether the escargot is tasty to her<sub>k</sub>*) will not have a true reading, for the same reasons as before.

Thus, the basic strategy used to deliver the right truth conditions for disagreement reports in §6.3 is not hostage to a relativist framework; it can work equally well in a contextualist framework. This suggests a further virtue of my approach to disagreement: my account is semantically neutral, in that it can be combined with either a relativist or a contextualist semantics for subjective expressions.<sup>42</sup>

It also suggests an important methodological upshot. Much of the literature has focused on *unembedded* occurrences of subjective expressions, as can be seen in the literature's tendency to emphasize discursive disagreement (§3). But insofar as we are interested in modeling doxastic disagreements, our focus should be on capturing belief reports embedding subjective expressions. Given the right semantic tools, both relativists and contextualists can deliver extensionally adequate accounts of the latter.

## 7 Conclusion

In this paper, I have sought a general account of subjective disagreement. I argued that the leading accounts either give us too little subjective disagreement or they give us too much. I went on to develop an alternative approach that fares better. My approach shares

---

<sup>42</sup>Which implementation—relativist or contextualist—should we prefer? Some authors seem to regard contextualism as the simpler and more common-sensical theory, which should be the default winner in the case of a tie. This is not my view. After all, the contextualist strategy developed here was only able to generate the right predictions by positing a special species of pronoun (PRO <sub>$\tau$</sub> ), as well as a special composition rule (IFA\*\*) that shifted the value of this pronoun. And so the resulting view is not in any obvious way simpler than the relativist alternative. I suspect that the choice between the contextualist and relativist implementations will ultimately rest on more general questions about the coherence of relativism, the relation between doxastic disagreement and discursive disagreement, and the conditions under which we retract assertions involving subjective expressions. I will not try to answer these questions here.

with the Simple Account the conviction that disagreement is a fundamentally doxastic phenomenon. However, rather than trying to cash out doxastic disagreement in terms of inconsistencies in the contents believed, my approach focuses on disagreement ascriptions, analyzing such ascriptions in terms of inconsistencies in the sentences used to characterize agents' beliefs.

The resulting account has a number of virtues. First, unlike other accounts, it gives us enough disagreement (it predicts e.g., the truth of (33)), without over-predicting disagreement among run-of-the-mill *de se* beliefs. Second, it explains why equivalent belief reports (e.g., (29) and (30)) sometimes generate different intuitions about disagreement. Third, it has the advantage of semantic neutrality: it can be implemented using either a relativist or a contextualist semantics for subjective expressions. Fourth, it retains many of the attractive features of the Simple Account of Disagreement. For example, it validates the following highly intuitive inference pattern:

- (42) a. A believes  $\phi$ .  
b. B believes  $\neg\phi$ .  $\Rightarrow$   
c. A and B disagree over whether  $\phi$ .

In light of these benefits, I think the approach to disagreement advanced in this paper is promising, and merits further development and exploration.

Is there any reason to be dissatisfied with my approach? A natural worry is that I have changed the topic of conversation. Our original question was not, 'What are the truth conditions for disagreement ascriptions?', but rather, 'What is disagreement?' Hasn't our original question been left unanswered? To put the worry another way: I have given a semantic analysis of disagreement *talk*, but I have not given a what we were really after, which is a metaphysical account of the nature of disagreement.

However, I think this worry is misplaced. My account does come with at least an implicit metaphysical picture. On this picture, there is an objective fact of the matter about what agents believe at various times. But there is no straightforward route from these doxastic facts to conclusions about disagreement. The only route is indirect: the doxastic facts, together with certain facts about our linguistic conventions (specifically, how we characterize beliefs in natural language) determine the truth or falsity of various belief ascriptions, which in turn determine the truth or falsity of various disagreement ascriptions. And that is all there is to say about disagreement. According to this view, the request for some further metaphysical account of disagreement is misguided: there are no objective disagreement facts over and above facts about the truth conditions for disagreement ascriptions.

An analogy may help. To take a by-now-familiar example, consider contextualist or relativist accounts of taste predicates. These accounts are typically accompanied by a metaphysical picture, according to which there are no facts about what is tasty *simpliciter*; rather, there are only facts of the form: [ $x$  enjoys the taste of  $y$  at time  $t$ ]. These facts, together with certain facts about our linguistic conventions, determine the truth or falsity of various tastiness ascriptions (that is, sentences of the form: ' $x$  is tasty'). But no further metaphysical account of tastiness can be given: there are no objective tastiness facts over and above facts about the truth conditions for tastiness ascriptions.

In case the reader finds this reply dissatisfying, it is worth observing that my account does enable us to give a metaphysical analysis of *objective* disagreement. An ascription of the form, 'A and B disagree over whether  $\phi$ ' reports an objective disagreement just in case the content of the complement clause ( $\phi$ ) can be modeled with a set of worlds (or a boring centered proposition). From this we can recover the simple story about objective disagreement with which we began: two parties objectively disagree over a set of worlds  $p$  iff one of them believes  $p$  and the other disbelieves  $p$ . Thus facts about objective agreement can be directly read off the doxastic facts, unlike the facts about subjective disagreement. This consequence can be used to further illuminate the differences between objective and subjective disagreement. One important difference is that whereas there are objective facts about objective disagreement, there are no objective facts about subjective disagreement, only objective facts about subjective disagreement ascriptions.

## References

- Derek Baker and Jack Woods. How expressivists can and should explain inconsistency. *Ethics*, 125: 391–424, 2015.
- Chris Barker. Negotiating taste. *Inquiry*, 56(2-3):240–257, 2013.
- Simon Blackburn. *Spreading the Word*. Clarendon Press, Oxford, 1984.
- Simon Blackburn. *Ruling Passions*. Clarendon Press, Oxford, 1998.
- Herman Cappelen and Josh Dever. *The Inessential Indexical*. Oxford University Press, Oxford, 2013.
- Herman Cappelen and John Hawthorne. *Relativism and Monadic Truth*. Oxford University Press, Oxford, 2009.
- Gennaro Chierchia. Anaphora and attitudes *De Se*. In Barsch, van Benthem, and van Emde Boas, editors, *Semantics and Contextual Expression*. Dordrecht, Foris, 1989.
- Dan López de Sa. Presuppositions of commonality: An indexical relativist account of disagreement. In García Carpintero and Kölbel, editors, *Relative Truth*. Oxford University Press, Oxford, 2008.
- Dan López de Sa. Expressing disagreement: A presuppositional indexical contextualist relativist account. *Erkenntnis*, 80:153–165, 2015.
- Keith DeRose. Single scoreboard semantics. *Philosophical Studies*, 119:1–21, 2004.

- Jamie Dreier. Relativism (and expressivism) and the problem of disagreement. *Philosophical Perspectives*, 23(1):79–110, 2009.
- Andy Egan. Secondary qualities and self-location. *Philosophy and Phenomenological Research*, 72: 97–119, 2006.
- Andy Egan. Epistemic modals, relativism, and assertion. *Philosophical Studies*, 133(1):1–22, 2007.
- Andy Egan. Disputing about taste. In Feldman and Warfield, editors, *Disagreement*. Oxford University Press, Oxford, 2010.
- Andy Egan. Relativist dispositional theories of value. *Southern Journal of Philosophy*, 50(4):557–582, 2012.
- Andy Egan. There's something funny about comedy: A case study in faultless disagreement. *Erkenntnis*, 79(1):73–100, 2014.
- Andy Egan, John Hawthorne, and Brian Weatherson. Epistemic modals in context. In Preyer and Peter, editors, *Contextualism in Philosophy*. Oxford University Press, Oxford, 2005.
- Hartry Field. Epistemology without metaphysics. *Philosophical Studies*, 143(2):249–290, 2009.
- Allan Gibbard. *Wise Choices, Apt Feelings*. Harvard University Press, Cambridge, MA, 1990.
- Allan Gibbard. *Thinking How to Live*. Harvard University Press, Cambridge, MA, 2003.
- Michael Glanzberg. Context, content, and relativism. *Philosophical Studies*, 136(1):1–29, 2007.
- Alex Grzankowski. Not all attitudes are propositional. *European Journal of Philosophy*, 3:374–391, 2012.
- Alex Grzankowski. Attitudes towards objects. *Noûs*, 50(2):314–328, 2016.
- Everett Hall. Stevenson on disagreement in attitude. *Ethics*, 58:51–56, 1947.
- Irene Heim and Angelika Kratzer. *Semantics in Generative Grammar*. Blackwell, Malden, MA, 1998.
- Jaakko Hintikka. *Knowledge and Belief*. Cornell University Press, 1962.
- Torfinn Thomesen Huvenes. Varieties of disagreement and predicates of taste. *Australasian Journal of Philosophy*, 90(1):167–181, 2012.
- Frank Jackson. Internal conflicts in desires and morals. *American Philosophical Quarterly*, 2:105–114, 1985.
- Justin Khoo. Modal disagreements. *Inquiry*, 5(1):1–24, 2015.
- Justin Khoo and Joshua Knobe. Moral disagreement and moral semantics. *Noûs*, forthcoming.
- Joshua Knobe and Seth Yalcin. Epistemic modals and context: Experimental data. *Semantics and Pragmatics*, 7(10):1–21, 2014.
- Max Kölbel. Faultless disagreement. *Proceedings of the Aristotelian Society*, 104:53–73, 2003.
- Peter Lasersohn. Context dependence, disagreement, and predicates of personal taste. *Linguistics and Philosophy*, 28:643–686, 2005.
- Peter Lasersohn. Non-world indices and assessment-sensitivity. *Inquiry*, 56(2–3):122–148, 2013.
- David Lewis. Attitudes *De Dicto* and *De Se*. *Philosophical Review*, 88:513–543, 1979.
- David Lewis. Dispositional theories of value. *Proceedings of the Aristotelian Society*, 63, 1989.
- John MacFarlane. Relativism and disagreement. *Philosophical Studies*, 132:17–31, 2007.
- John MacFarlane. Epistemic modals are assessment sensitive. In Egan and Weatherson, editors, *Epistemic Modality*. Oxford University Press, Oxford, 2011.
- John MacFarlane. *Assessment-Sensitivity: Relative Truth and its Applications*. Oxford University Press, Oxford, 2014.

- Teresa Marques and Manuel García-Carpintero. Disagreement about taste: Commonality presuppositions and coordination. *Australasian Journal of Philosophy*, 92(4):701–723, 2014.
- Jerry Morgan. On the criterion of NP deletion. *CLS*, 6:380–389, 1970.
- Dilip Ninan. *De Se* attitudes: Ascription and communication. *Philosophy Compass*, 5(7):551–567, 2010.
- Hazel Pearson. A judge-free semantics for predicates of personal taste. *Journal of Semantics*, 30: 103–154, 2013.
- John Perry. The problem of the essential indexical. *Noûs*, 13:3–32, 1979.
- David Plunkett and Timothy Sundell. Disagreement and the semantics of normative and evaluative terms. *Philosophers' Imprint*, 13:1–37, 2013.
- Mark Richard. Contextualism and relativism. *Philosophical Studies*, 119(1-2):215–242, 2004.
- Mark Richard. *When Truth Gives Out*. Oxford University Press, New York, 2008.
- Mark Richard. Relativistic content and disagreement. *Philosophical Studies*, 156(3):421–431, 2011.
- Michael Ridge. Disagreement. *Philosophy and Phenomenological Research*, 86(1):41–63, 2013.
- Jacob Ross and Mark Schroeder. Reversibility or disagreement. *Mind*, 122:43–84, 2013.
- Jonathan Schaffer. Perspective in taste predicates and epistemic modals. In Egan and Weatherston, editors, *Epistemic Modality*. Oxford University Press, Oxford, 2011.
- Philippe Schlenker. A plea for monsters. *Linguistics and Philosophy*, 26(1):29–120, 2003.
- Robert K. Shope. The conditional fallacy in contemporary philosophy. *The Journal of Philosophy*, 75 (8):397–413, 1978.
- Alex Silk. Normative language in context. In Shafer-Landau, editor, *Oxford Studies in Metaethics*, volume 12. Oxford University Press, New York, forthcoming.
- Neil Sinhababu. Advantages of propositionalism. *Pacific Philosophical Quarterly*, 96(1):165–180, 2015.
- Tamina Stephenson. Judge dependence, epistemic modals, and predicates of personal taste. *Linguistics and Philosophy*, 30(4):487–525, 2007a.
- Tamina Stephenson. Indicative conditionals have relative truth conditions. *Proceedings of the Annual Meeting of the Chicago Linguistics Society*, 43(1):231–242, 2007b.
- Tamina Stephenson. *Towards a Theory of Subjective Meaning*. Ph.D. Thesis, MIT, 2007c.
- Tamina Stephenson. Control in centered worlds. *Journal of Semantics*, 27(4):409–436, 2010.
- C.L. Stevenson. *Ethics and Language*. Yale University Press, New Haven, 1944.
- C.L. Stevenson. *Facts and Values*. Greenwood Press, Westport, CT, 1963.
- Isidora Stojanovic. Talking about taste: Disagreement, implicit arguments, and relative truth. *Linguistics and Philosophy*, 30:691–706, 2007.
- Timothy Sundell. Disagreement about taste. *Philosophical Studies*, 39(3):359–396, 2011.
- Eric Swanson. *Interactions with Context*. MIT PhD dissertation, 2006.
- Malte Willer. Dynamics of epistemic modality. *Philosophical Review*, 122(1):44–92, 2013.
- Seth Yalcin. Nonfactualism about epistemic modality. In Egan and Weatherston, editors, *Epistemic Modality*. Oxford University Press, Oxford, 2011.

## Appendix A. Formal Derivations

This appendix uses the semantic clauses from §6 to derive truth conditions for *de se* belief ascriptions. §A.1 uses the semantics for Chierchia's abstraction operator  $O$  to derive Lewisian truth conditions for ordinary *de se* belief ascriptions such as (11'). §A.2 combines this approach with Stephenson's relativist treatment (§6.3) to derive the result that taste belief ascriptions lacking overt pronouns are equivalent to (the *de se* readings of) taste belief ascriptions involving overt pronouns. Finally, §A.3 shows that the same results can be derived using the contextualist alternative developed in §6.4.

### A.1. Truth conditions for attitude ascriptions involving overt pronouns

Here we derive the Lewisian truth conditions for (11') ( $\text{Kim}_i \text{ believes } [O_i. \text{she}_i \text{ is hungry}]$ ). Our lexical entries for *she*,  $O$ , and *believes* are taken from §6.1. Our composition rules are standard versions of Function Application (FA) and Intensional Function Application (IFA) (Heim and Kratzer 1998):

**FA** If  $\alpha$  is a complex expression formed by combining two expressions  $\beta$  and  $\gamma$ ,

and  $[\![\gamma]\!]^{c,g,w}$  is in the domain of  $[\![\beta]\!]^{c,g,w}$ , then  $[\![\alpha]\!]^{c,g,w} = [\![\beta]\!]^{c,g,w}([\![\gamma]\!]^{c,g,w})$ .

**IFA** If  $\alpha$  is a complex expression formed by combining two expressions  $\beta$  and  $\gamma$ ,

and  $[\![\lambda w'_s. [\![\gamma]\!]^{c,g,w'}]\!]^{c,g,w}$  is in the domain of  $[\![\beta]\!]^{c,g,w}$ , then  $[\![\alpha]\!]^{c,g,w} = [\![\beta]\!]^{c,g,w}([\![\lambda w'_s. [\![\gamma]\!]^{c,g,w'}]\!])$ .

By combining these ingredients, we predict that (11') is true iff Kim believes HUNGRY:

$$\begin{aligned}
 & ([\!(\text{Kim}_i \text{ believes } [CP O_i. \text{she}_i \text{ is hungry}])\!]^{c,g,w} = 1 \text{ iff} \\
 & \quad [\![\text{believes } [O_i \text{ she}_i \text{ is hungry}]]\!]^{c,g,w}([\![\text{Kim}_i]\!]^{c,g,w}) = 1 \text{ iff} && \text{by FA} \\
 & \quad [\![\text{believes}]\!]^{c,g,w}(\lambda w'. [\![O_i \text{ she}_i \text{ is hungry}]\!]^{c,g,w'})([\![\text{Kim}_i]\!]^{c,g,w}) = 1 \text{ iff} && \text{by IFA} \\
 & \quad \forall \langle w'', x \rangle \in \text{Dox}(w, \text{Kim}): \lambda w'. [\![O_i \text{ she}_i \text{ is hungry}]\!]^{c,g,w'}(w'')(x) = 1 \text{ iff} && \text{by } \text{believes} \\
 & \quad \forall \langle w'', x \rangle \in \text{Dox}(w, \text{Kim}): \lambda w'. \lambda x'. [\![\text{she}_i \text{ is hungry}]\!]^{c,g[x'/i],w'}(w'')(x) = 1 \text{ iff} && \text{by } O_i \\
 & \quad \forall \langle w'', x \rangle \in \text{Dox}(w, \text{Kim}): \lambda w'. \lambda x'. x' \text{ is hungry at } w'(w'')(x) = 1 \text{ iff} && \text{by FA, } \text{she} \\
 & \quad \forall \langle w'', x \rangle \in \text{Dox}(w, \text{Kim}): x \text{ is hungry at } w''. && \text{by } \lambda\text{-conversion}
 \end{aligned}$$

### A.2 Relativist Treatment of Taste Beliefs

We now show that combining Chierchia's approach with Stephenson's treatment of taste predicates predicts that taste beliefs ascriptions involving overt pronouns have *de se* truth conditions that are equivalent to those of taste beliefs ascriptions lacking such pronouns. In particular, we derive the result that (31) ( $\text{Ava} \text{ believes } [\text{the escargot is tasty } PRO_J]$ ) and (32) ( $\text{Ava}_i \text{ believes } [O_i \text{ the escargot is tasty to her}_i]$ ) are both true iff Ava believes TASTY.

To derive truth conditions for (31), we use Stephenson's lexical entries for *tasty* and  $PRO_J$ , as well as her modified version of Intensional Function Application (IFA\*) (§6.3):

$$\begin{aligned}
 & [\text{Ava believes } [CP \text{ the escargot is tasty } PRO_J]]^{c,g,w,j} = 1 \text{ iff} \\
 & [\text{believes } [\text{the escargot is tasty } PRO_J]]^{c,g,w,j}([\text{Ava}]^{c,g,w,j}) = 1 \text{ iff} && \text{by FA} \\
 & [\text{believes}]^{c,g,w,j}(\lambda w'.\lambda j'.[\text{the escargot is tasty } PRO_J]^{c,g,w',j'})([\text{Ava}]^{c,g,w,j}) = 1 \text{ iff} && \text{by IFA*} \\
 & \forall \langle w'', x \rangle \in \text{Dox}(w, \text{Ava}) : \\
 & \quad \lambda w'.\lambda j'.[\text{the escargot is tasty } PRO_J]^{c,g,w',j'}(w'')(x) = 1 \text{ iff} && \text{by } \text{believes} \\
 & \forall \langle w'', x \rangle \in \text{Dox}(w, \text{Ava}) : \\
 & \quad \lambda w'.\lambda j'. \text{ the escargot is tasty to } j' \text{ at } w'(w'')(x) = 1 \text{ iff} && \text{by } \text{tasty}, PRO_J \\
 & \forall \langle w', x' \rangle \in \text{Dox}(w, \text{Ava}) : \text{ the escargot is tasty to } x' \text{ in } w'. && \text{by } \lambda\text{-conversion}
 \end{aligned}$$

To derive truth conditions for (32), we rely on Stephenon's lexical entry for *tasty*, together with our semantics for O:

$$\begin{aligned}
 & [\text{Ava}_i \text{ believes } [O_i \text{ the escargot is tasty to her}_i]]^{c,g,w,j} = 1 \text{ iff} \\
 & [\text{believes } [O_i \text{ escargot is tasty to her}_i]]^{c,g,w,j}([\text{Ava}_i]^{c,g,w,j}) = 1 \text{ iff} && \text{by FA} \\
 & [\text{believes}]^{c,g,w,j}(\lambda w'.[O_i \text{ escargot is tasty to her}_i]^{c,g[x/i],w',j})([\text{Ava}_i]^{c,g,w,j}) = 1 \text{ iff} && \text{by IFA} \\
 & \forall \langle w'', x' \rangle \in \text{Dox}(w, \text{Ava}) : \\
 & \quad \lambda w'.\lambda x.[\text{the escargot is tasty to her}_i]^{c,g[x/i],w',j}(w''), (x') = 1 \text{ iff} && \text{by } \text{believes}, O_i \\
 & \forall \langle w'', x' \rangle \in \text{Dox}(w, \text{Ava}) : \\
 & \quad \lambda w'.\lambda x. \text{ the escargot tastes good to } x \text{ at } w'(w''), (x') = 1 \text{ iff} && \text{by } \text{tasty} \\
 & \forall \langle w', x' \rangle \in \text{Dox}(w, \text{Ava}) : \text{ the escargot is tasty to } x' \text{ in } w'. && \text{by } \lambda\text{-conversion}
 \end{aligned}$$

### A.3 Contextualist Treatment of Taste Beliefs

Here we show that our contextualist alternative also predicts that taste belief ascriptions have *de se* truth conditions. In particular, it predicts that (39) (*Ava believes [the escargot is tasty*  $PRO_\tau$ *]*) is true iff Ava believes TASTY, rendering it equivalent to (32). Our semantics for *tasty* and  $PRO_\tau$  are taken from §6.4, as is our modified composition rule IFA\*\*:

$$\begin{aligned}
 & [\text{Ava believes } [CP \text{ the escargot is tasty } PRO_\tau]]^{c,g,w} = 1 \text{ iff} \\
 & [\text{believes } [\text{the escargot is tasty } PRO_\tau]]^{c,g,w}([\text{Ava}]^{c,g,w}) = 1 \text{ iff} && \text{by FA} \\
 & [\text{believes}]^{c,g,w}(\lambda w'_s.\lambda x_e.[\text{the escargot is tasty } PRO_\tau]^{c,g[x/\tau],w'})(\text{Ava}) = 1 \text{ iff} && \text{by IFA**} \\
 & \forall \langle w'', x' \rangle \in \text{Dox}(w, \text{Ava}) : \\
 & \quad \lambda w'.\lambda x.[\text{the escargot is tasty } PRO_\tau]^{c,g[x/\tau],w'}(w'')(x') = 1 \text{ iff} && \text{by } \text{believes} \\
 & \forall \langle w'', x' \rangle \in \text{Dox}(w, \text{Ava}) : \\
 & \quad \lambda w'.\lambda x. \text{ the escargot is tasty to } x \text{ at } w'(w'')(x') = 1 \text{ iff} && \text{by } \text{tasty}, PRO_\tau \\
 & \forall \langle w', x \rangle \in \text{Dox}(w, \text{Ava}) : \text{ the escargot is tasty to } x \text{ in } w'. && \text{by } \lambda\text{-conversion}
 \end{aligned}$$