TWISTED WAYS TO SPEAK OUR MINDS, OR WAYS TO SPEAK OUR TWISTED MINDS?

Luis Rosa
University of Cologne

Forthcoming in *Epistemology of Conversation* (Springer Series), edited by Waldomiro J. Silva Filho.

There are many ways in which a speaker can confuse their audience. In this paper, I will focus on one such way, namely, a way of talking that seems to manifest a cross-level kind of cognitive dissonance on the part of the speaker. The goal of the paper is to explain why such ways of talking sound so twisted. The explanation is two-pronged, since their twisted nature may come *either* from the very mental states that the speaker thereby makes manifest, *or* from how the speaker chooses to express themselves (even if there is nothing wrong with their mental states). So-called 'Moore-paradoxical' utterances are but one example of the phenomenon, and the explanation of what is wrong about them is subsumed under a more general explanation here—one that captures also the twisted-ness of utterances whereby questions are raised or intentions expressed.

1 Introduction

Cognitive dissonance is a phenomenon studied not only by psychologists, but also by philosophers.[1] There is cognitive dissonance and then there are its purported manifestations or signs, which include certain *speech acts.* In philosophy, much attention has been paid to the latter under the name of 'Moore-paradoxical' utterances, such as a speaker's utterance of 'I am tired, but I don't believe I am'.[2]

Relatedly, philosophers have also tried to determine what, if anything, makes *akratic* mental states normatively subpar.[3] A typical example of that is a mental state where one intends to do something while believing one ought not to do it. The speech acts whereby such mental states are made manifest also signal cognitive dissonance. When a

[1] See Festinger (1962), Elliot and Devine (1994) for how the notion of cognitive dissonance is used in psychology.

[2] See Moore (1993) and, for example, de Almeida (2001) and Williams (2015).

[3] See for example Worsnip (2018), Lasonen-Aarnio (2020) and Rosa (2022) for some of the discussion.

speaker utters ‘I want to buy a new mattress, but I shouldn’t’, for example, she leaves her interlocutors thinking that she is somewhat at odds with herself.

In this paper, I am going to deal with a large class of utterances that make cognitive dissonance manifest. A mark of the utterances belonging to that class is that they all feature a step from lower- to higher-order speech: the speaker first says something (lower-order talk), and then they say something about their standing with respect to what they said before (higher-order talk). The typical Moore-paradoxical utterances are but one example of the more general phenomenon under scrutiny here, and so are utterances that seem to indicate the obtaining of akratic mental states. Those are both special cases of a more general phenomenon.

Any utterance from that large class, I claim, is either a *twisted way to speak one’s mind* or a *way to speak one’s twisted mind.* In the latter case, the speaker’s mind harbors some kind of error or defect, and the way of talking inherits its twisted quality from the mental states it is a manifestation of. This dichotomy will reappear below. The utterances that belong to the target class include utterances of the following types (where ‘*p?*’ is an interrogative rendering of the declarative ‘*p*’—more on this below):

(1) *p*, but I don’t know that *p*.

(2) *p*, but it is irrational for me to believe that *p*.

(3) *p?* Though I shouldn’t be in doubt about whether *p*.

(4) I will be there! Though it is irrational for me to intend to be there.

(5) *p*, but I know that $\neg p$.

(6) *p*, but it is rational for me to be in doubt about whether *p*.

(7) *p?* Though I know whether *p*.

(8) *p?* Though it is rational for me to believe that *p*.

(9) I will be there! Though I should want to be somewhere else.

(10) *p*, but I don’t believe that *p*.

(11) *p?* Though I am not in doubt about whether *p*.

(12) I will be there! Though I do not intend to.

(13) *p*, but I believe that $\neg p$.

(14) *p*, but I am in doubt about whether *p*.

(15) *p*, but probably $\neg p$.

(16) *p?* Though I believe that *p*.

(17) *p?* Though surely *p*.

(18) I will be there! Though I intend to be somewhere else.

The paper will further divide that large class of utterances into two subclasses. The utterances that belong to the first subclass, which I call 'disapprovals', are such that their higher-order bit involves the use of *normative* expressions or expressions of *evaluation* or *appraisal* (e.g., 'should', 'rational', 'know', etc.). These include the utterances of sentences of forms (1)–(9) from above. One concrete example of a disapproval is the utterance of 'God exists, but I shouldn't believe that God exists'.

The utterances that belong to the second subclass, which I call 'disavowals', are such that their higher-order bit involves *psychological* vocabulary only (e.g., the verbs 'believe', 'doubt', 'be sure', etc.) instead of normative expressions or expressions of evaluation/appraisal. These include the utterances of sentences of forms (10)–(18) from above. One concrete example of a disavowal is a speaker's utterance of 'I have no doubt that God exists, but does he?'.

The task of the paper will be to explain what is wrong or twisted about all of these utterances. The explanation of what is wrong about disapprovals, however, is not exactly the same as the explanation of what is wrong about disavowals. Some important commonalities aside, they are twisted in different ways. It is noted along the way that certain explanations about the wrongness or oddity of Moore-paradoxical utterances offered in the literature lack the level of generality that is needed to explain why disapprovals and disavowals in general sound so bad.

2 Expressing and saying

Before presenting the relevant explanations, however, I need to make some of my assumptions explicit.

I will use a notion of speaker's *expressing* an attitude/its absence thereof through an utterance. I take that relation to be a three-place relation between a speaker, a mental state and an uttered sentence (a sentence *token*, including declarative, interrogative and imperative sentence tokens). Examples of that relation include a speaker's expressing her belief that *some snakes are poisonous* by uttering 'Some snakes are poisonous', a speaker's expressing her intention to be some place by uttering 'I will be there!', and a speaker's expressing her state of doubt or uncertainty as to whether *there are intelligent aliens* by uttering the interrogative sentence 'Are there intelligent aliens?'.

The notion of one's expressing one's attitudes/mental states through an utterance is of course crucial to expressivist theories of assent and dissent patterns of different natural language constructions.[4] Expressivists often take specific types of adjectives or verbs to play the role of helping convey some aspect of the speaker's mental life, as opposed to making a semantic contribution to how the speaker says the world is like (the world beside their mind, that is—if they do that at all). For example, the Bayesian form of expressivism put forward by Yalcin (2021, p. 125) takes it that in uttering:

(0) Allan is probably in the office,

a speaker expresses their credal state with respect to whether *Allan is in the office*, without literally *saying that* they are in that credal state. The speaker thereby makes manifest her credal state of being more confident that *Allan is in the office* than that *Allan is not in the office*, as opposed to manifesting outright belief towards a propositional content to the effect that it is more probable that *Allan is in the office than not.*

The explanation that I will offer below is *consistent* with such forms of expressivism, but it is not *committed* to them. I will assume, for example, that one can express one's belief that *p* not only by uttering '*p*', but

[4] See for example Gibbard (1990) and Schroeder (2008). For a recent expressivist proposal to diagnose Moore-paradoxical phenomena, see Freitag and Yolcu (2021).

also by uttering ‘I believe that *p*’, and that one can express one’s certainty that *p* by uttering ‘Certainly *p*’, etc. (same for similar verbs/adverbs). Expression in the present sense is a relatively *cheap* phenomenon, in that there are many different linguistic means through which one can express features of one’s mental states, which includes the utterance of sentences featuring propositional attitude-verbs and adverbs as a particular case. But that does not entail that the only function, or even the main function of those verbs and adverbs is to express mental states or attitudes when combined with first-person pronouns.

As already hinted at, *expressing* an attitude (or its absence thereof) is to be contrasted with *saying that* one holds (doesn’t hold) that attitude. In order to say that one holds (doesn’t hold) a given attitude, one must utter a sentence whose semantic value is the proposition that *one holds* (*doesn’t hold*) *that attitude.* To say that one holds (doesn’t hold) a given attitude is to impart that information via the utterance of a sentence that carries that very piece of information (the sentence itself is a vehicle that carries that information—think of the proposition that is the semantic value of the sentence as the information it carries).

In order for one to say that one believes in miracles, for example, one must use the verb ‘believe’ or some synonym/translation of it. One cannot literally say that one believes that it is raining by uttering ‘It is raining’. But one can express one’s belief that *it is raining* by uttering ‘It is raining’. In contrast to the relation of saying, the expression of a mental state is much less tethered to linguistic form—it floats free across a multitude of linguistic expressions. One doesn’t *have* to use the verb ‘believe’ or some synonym/translation of it in order to express one’s beliefs, though one has to use such expressions in order to literally say that one has those beliefs.[5]

(For obvious reasons, the relation of expressing an attitude/its absence thereof through an utterance is to be distinguished from the expression relation that holds between a sentence and a proposition, as when we say that ‘*p*’ expresses the proposition that *p*. To avoid confu-

[5] This is compatible, however, with the claim that, when it comes to speech acts (ignoring other kinds of actions), one does have to use certain kinds of verbs and adverbs in order to express other aspects of one’s mental life—for example, that one has to use some device like ‘probably’ in order to express one’s credences.

sion, I only use the verb ‘express’ in the former sense here).[6]

What is it for a speaker to express her mental states through an utterance? That is a difficult question, but minimally it involves this: in making the utterance, the speaker signals that she is in the relevant mental state to potential hearers—her utterance serves as *evidence* that she is in that state—without necessarily *saying that* she is in it. But this is not the place to try harder than that to explicate the relevant notion of expression.

3 Disapprovals

Every utterance that I will examine here is thought of as part of a single speech act, even where the sentences uttered might suggest that separate speech acts were performed (at different contexts). I adopt a convenient structuring of the target sentences, where their lower-order bit appears first and their higher-order bit appears second, reading from left to right, as in the items from the list (1)–(18) above. Thus, ‘I know whether it is raining, but is it raining?’ is paraphrased through ‘Is it raining? Though I know whether it is raining’.

I use ‘*p*’ as a placeholder for declarative sentences, and ‘*p?*’ as a placeholder for interrogative sentences. ‘*p?*’ is the interrogative rendering of the declarative ‘*p*’. For example, ‘Is it raining?’ is the interrogative rendering of the declarative sentence ‘It is raining’.

The first group of utterances to be investigated here is the group of disapprovals. There are *direct* and *indirect* disapprovals. Here are some examples of direct disapprovals—think of a speaker uttering any of the following:

(1) *p*, but I don’t know that *p*.

(2) *p*, but it is irrational for me to believe that *p*.

(3) *p?* Though I shouldn’t be in doubt about whether *p*.

(4) I will be there! Though it is irrational for me to intend to be there.

[6] See also Bar-On (2015) on this point.

Here is a mark of direct disapprovals: the sentence used in their higher-order bit describes the mental state that the speaker expresses through their first-order bit as having some negative (deontic or epistemic) status. In the second bit of the utterance, that is, the speaker disapproves of the attitude that they have expressed in the first bit.

Consider (2), for example. When the speaker asserts that *p* by uttering '*p*', they thereby express belief towards *p*—but then they go on and say that it is irrational for them to hold that very belief. In (3), the speaker asks whether *p* is the case by uttering '*p?*', and they thereby express a state of being in doubt about whether *p* is the case or not—but then they go on and say that they shouldn't be in that very state. Similarly, in the second bit of (4) the speaker says that their intention to be somewhere, which they have expressed by uttering 'I will be there!' in the first bit, is not a rational intention for them to have.

Indirect disapprovals differ from direct disapprovals in that the second bit of the former ones doesn't *directly* criticize the attitude expressed in the first bit—though some criticism of the attitude expressed in the first bit *follows from* what is said in the second bit. For example, think of a speaker uttering any of the following:

(5) *p*, but I know that $\neg p$.

(6) *p*, but it is rational for me to be in doubt about whether *p*.

(7) *p?* Though I know whether *p*.

(8) *p?* Though it is rational for me to believe that *p*.

(9) I will be there! Though I should want to be somewhere else.

Contrast (1) and (5), for example. In (1), the second bit directly criticizes the attitude that the speaker has expressed in the first bit (a belief that is not knowledge). In contrast, the second bit of (5) makes no such direct criticism, though it *follows from* what is said in it that the belief expressed in the first bit is not knowledge: if one knows that $\neg p$ then one does not know that *p*.

Similarly, it follows from what is said in the second bit of both (7) and (8) that the attitude of being in doubt about whether *p*, which the speaker has expressed through their first bit, is not a 100% normatively on the clear. In the case of (7), in at least one sense of 'should', it follows from the fact that one knows whether *p* that one shouldn't be in doubt about whether *p* is the case. In the case of (8), it follows from the fact that it is rational for one to believe that *p* that it is not rational, or at least not perfectly rational, for one to be in doubt about whether *p* is the case.

Some philosophers may want to dispute some of these claims—but I won't try to fix all the holes one can try to poke at the claim that (7)/(8) are disapprovals, as much as the other examples are, in that they also seem to make manifest a kind of dissonance between the speaker's attitudes and her own assessment of those attitudes.[7] If the reader doesn't want to lump these examples together with the other ones, then they are free to just think of the other ones under the label of 'disapprovals'. Still a big class of utterances is left, and many other examples besides the ones offered so far can be fleshed out.

Now why are (1)–(9) such twisted ways of talking?

4 Why disapprovals sound so twisted

Any of the intensional attitudes mentioned above can be normatively subpar, or be somehow at fault or less than ideal, given the kind of attitude that it is. In holding the attitude, the cognizer fails to abide by some norm for that attitude, or that attitude falls short of its axiological (epistemic or practical) ideals.

For example, there is a norm of epistemic rationality according to which one should believe that *p* only if one's evidence supports *p*.[8] When one believes that *p* on the basis of insufficient evidence, one's belief is thereby normatively subpar. Another epistemic norm for belief says that one should believe that *p* only if one knows that *p*.[9] Or, at the very least, belief *at its best* is knowledge. When it isn't, it falls short of

[7] Those who agree with Lewis (1982) and Stalnaker (1984) that our minds admit of different *fragments* (ways of framing things that facilitate access to different bits of information) are invited to relativize ascriptions of attitudes/expressed attitudes to the very same fragment. See Borgoni, Kindermann and Onofri (2021) for a recent volume on the issue of fragmentation and how it bears on a number of epistemological issues.

[8] See e.g. Feldman (2000) and Williamson (2000, Ch. 8).

[9] See e.g. Smithies (2012), Littlejohn (2013).

that epistemic ideal, or it is axiologically subpar (it could be better).

And similarly for other intensional attitudes such as that of being in doubt about whether something is the case and that of intending that something is the case. Each of these attitudes have their own norms and axiological ideals.[10]

Now consider a speaker's sincere utterance of (1), again:

(1) *p*, but I don't know that *p*.

In uttering the first bit, the speaker *expresses* an attitude of belief that *p*. In uttering the second bit, the speaker *says that* they don't know that *p*. The reason why a speaker's utterance of (1) sounds problematic, we might think, is that *if what the speaker said* through the second bit is true, then *what the speaker expressed* through the first bit is normatively/axiologically subpar.

In sincerely uttering (1), then, the speaker is thus guaranteed to make some kind of error: either they spoke falsely in the second bit (they *do* know that *p*, even though they said they don't), or they spoke truly in the second bit, in which case the attitude that they have expressed through the first bit is normatively subpar. So *no matter how the world is like*, either the speaker's speech act is defective (in that they have asserted a falsehood) or their speech act is not that defective, but the belief that they have expressed is defective.[11]

Think of it as follows. The speaker has uttered (1) and you have heard them. Can you take their word for it? Suppose you do. So now, based on the second bit of their utterance, you conclude that they don't know that *p*. And, based on the first bit of their utterance, you conclude that they believe that *p*. Putting those two things together, then, you conclude further that they hold a belief that falls short of the ideal of knowledge.

If the speaker is right—right in what they're saying in the second bit—then the speaker is making some kind of mistake—a mistake in the

[10] E.g., Shah (2008) and also McHugh and Way (2018) adopt permissibility as a standard of correctness or fittingness for intention.

[11] I cannot quite exactly determine how this diagnosis regarding (1) is to the many purported solutions of Moore's paradox—see Green and Williams (2007) for a sample of that. It is in any case a diagnosis that generalizes to the other forms of disapproval, including those whereby the speaker asks questions and makes promises.

attitude that they have expressed in the first bit, that is, the mistake of believing without knowing. You cannot take their word for it without finding some grounds for criticizing them (consider the criticism: 'But you don't know that!'). Otherwise, the speaker is wrong in saying what they said in the second bit. Either way, a mistake has been made by the speaker. We reach a similar conclusion with respect to a speaker's sincere utterance of (5), with only one extra step from the assumption that the speaker knows that $\neg p$ to the conclusion that they don't know that p.

Similarly, consider a speaker's sincere utterance of (3), again:

(3) *p?* Though I shouldn't be in doubt about whether p.

In uttering the first bit, the speaker expresses a state of doubt about whether p is the case. In uttering the second bit, the speaker says that they shouldn't be in doubt about whether p. The reason why a speaker's utterance of (3) sounds problematic, we might think, is that *if what the speaker said* through the second bit is true, then *what the speaker expressed* through the first bit is normatively subpar. Either that or the speaker didn't say the truth through the second bit. No matter how the world is like, again, either the speaker's speech act is defective (in that they have asserted a falsehood) or the speech act is not that defective—but then the state of uncertainty that they have expressed through the first bit is defective (they shouldn't be in it).

We reach a similar conclusion with respect to a speaker's sincere utterance of (7), with only one extra step from the assumption that the speaker knows whether p is the case to the conclusion that they shouldn't be uncertain about whether p is the case.

Summing up, regarding any disapproval: either *what the speaker says* through their second bit is false or, if it isn't, then *what the speaker expresses* through their first bit is normatively/axiologically subpar. A hearer cannot take their word for it without finding some grounds to criticize the attitudes they have expressed through their speech act. That conclusion is held, of course, under the assumption that the speaker is making a *sincere* utterance. Otherwise, if their utterance is not sincere, in that they do not hold the attitudes that they thereby express, then their speech act is problematic on that very count.

And that is why disapprovals sound so twisted. Either they are twisted ways to speak one's mind—because they are insincere, or sincere but inaccurate—or, assuming sincerity and accuracy, they are ways of speaking one's twisted mind—because the mental states thereby expressed are guaranteed to be normatively/axiologically subpar.

Notice that this explanation has the degree of generality that is needed to explain what is wrong about *all* of (1)–(9). In this it contrasts with other explanations of the absurdity of Moore-paradoxical utterances in the literature—for example, ones using the unknowability of (1) plus a knowledge norm of assertion (as in Williamson 2000, Ch. 11). Sure enough, it is impossible for one to know that p and know at the same time that *one doesn't know that p.* But how does that explain what is wrong about (3), for example? (A state of doubt is not a state of belief, therefore it is not the kind of state that is in the game to become knowledge).

Similarly, Whitcomb (2017) explains the incoherence of what he calls 'Moore-paradoxical questions', such as a speaker's utterance of 'I know it is snowing, but is it snowing?' and 'Am I the only omniscient being?' by appealing to a constitutive norm of inquiry, namely, that one should inquire into a given question only if one doesn't already know what the true answer to that question is. But how does that very explanation tell us what is wrong with (1)? Some explanations explain the twisted-ness of (1) without explaining the twisted-ness of (3), others explain the twisted-ness of (3) without explaining the twisted-ness of (1). And here I am offering an explanation that explains the twisted-ness of all of (1)–(9) at the same time.

5 The mental counterparts of disapprovals

What about the 'purely mental' counterparts of (1)–(9)? That does *not* mean: Why is it problematic for one to *believe* each of (1)–(9)? So understood, the question is ill-formed. There isn't such a thing, for example, as believing that *p? Though I know whether p.* Since one cannot believe a question, one cannot believe the conjunction or concatenation of a question and a proposition.

A better sense of the initial question is made as follows. Let '(n)' be a variable ranging over all of (1)-(9). There seems to be a problem with a total intensional state that is constituted by both, the attitude that is *expressed* through the first bit of an utterance of (n) and the attitude that is *expressed* through the second bit of an utterance of (n). For example, regarding (1), when the speaker utters '*p*' they thereby express belief that *p*, and when they utter 'I don't know that *p*' they thereby express belief that *they don't know that p*. Now the question is: what is wrong with a person's doxastic state when they believe that *p* and they believe at the same time that *they don't know that p*? And, regarding (3), what is wrong with a person's doxastic state when they are in doubt about whether *p* and they believe at the same time that they *shouldn't be in doubt about whether p*? And, regarding (4), what is wrong with a person's intensional state when they intend to be somewhere and they believe at the same time that it is irrational for them to intend to be there? Etc.

Now the twisted ways of *talking*—the twisted utterances—are out of the picture, and we just have to determine what is twisted about the mind that harbors the relevant combinations of intensional attitudes. But the considerations from above already provide us with a ready answer to the question of what makes them so: it is *impossible* for all of the attitudes of such combinations to simultaneously abide by their respective norms or live up to their axiological ideals (norms and ideals for belief, intension, doubt, etc.). At least one of the relevant attitudes is *guaranteed* to be normatively/axiologically subpar, assuming that the other one isn't.

Consider for example believing that *p* while at the same time believing that *one doesn't know that p* (doxastic analog of (1)). Assume that the latter belief is *not* in any way normatively or axiologically subpar. Then one knows that *one doesn't know that p*. So one doesn't know that *p*, because knowledge is factive. But that entails that their belief that *p* is axiologically subpar (it doesn't constitute knowledge). Or consider being in doubt about whether *p* while at the same time believing that *one shouldn't be in doubt about whether p* (doxastic analog of (3)). Assume that the latter belief is not axiologically subpar. Then one knows that *one shouldn't be in doubt about whether p*. So one shouldn't be in doubt about whether *p*. But that entails that their state of doubt about whether *p* is normatively subpar (they shouldn't be in that state).

And so on. The same kind of explanation can be given about why any of the doxastic analogues of each of (1)–(9) are problematic.

6 Disavowals

In §4 I have offered an explanation of why disapprovals sound so twisted. They are either *twisted ways to speak our minds*—because they are insincere, or rather sincere but inaccurate—or they are *ways of speaking our twisted minds*—because the mental states they express are guaranteed to be normatively/axiologically subpar. Now it is time to tackle disavowals, the second big group of twisted utterances investigated here.

As I mentioned in §1, the explanation of what is so twisted about disapprovals is not exactly the same as the explanation of what is so twisted about disavowals. Similarly to disapprovals, however, disavowals also come in *direct* and *indirect* versions. Here are some examples of direct disavowals—think of a speaker sincerely uttering:

(10) p, but I don't believe that p.

(11) p? Though I am not in doubt about whether p.

(12) I will be there! Though I do not intend to.

A speaker making any of these utterances in the context of a conversation is bound to leave their interlocutors confused. Does the speaker of (10) believe that p or not? Is the speaker of (11) in doubt about whether p or not? Etc. They all sound like they are a bit mixed, ambivalent.

Here is a mark of direct disavowals: the sentence used in the second bit say that the speaker does not hold the attitude that they have expressed through the first bit. Another way of putting it: the sentence used in the second bit *denies that* the speaker holds the attitude that they have expressed through the first bit. And it is exactly here where the difference between direct and indirect disavowals lies, for the sentence used in the second bit of *indirect* disavowals does *not* itself deny that the speaker has the attitude that they have expressed through the first bit. Consider some examples of indirect disavowals:

(13) p, but I believe that $\neg p$.

(14) p, but I am in doubt about whether p.

(15) p, but probably $\neg p$.

(16) *p?* Though I believe that *p*.

(17) *p?* Though surely *p*.

(18) I will be there! Though I intend to be somewhere else.

For example, neither (13) nor (14) directly deny that the speaker believes that *p*, which is the attitude they have expressed by asserting that *p* in their first bit—though it might be argued that it *follows from* what is said in their second bit that the speaker does not believe that *p* (in case it is *impossible* for one to believe that *p* and believe that $\neg p$ at the same time, impossible for one to believe that *p* and be in doubt about whether *p* at the same time). This particular issue need not be addressed here, however, for the explanation that I will offer for the twisted-ness of indirect disavowals is not committed to the claim that such entailment relations in fact hold.

Now notice that there is an important difference between (15) and (17), on the one hand, and the remaining disavowals from that list, on the other. (15) and (17) feature *adverbs*, as opposed to intensional attitude *verbs* that connect the grammatical subject 'I' to some complex construction (such as a declarative sentence), as in the other examples of disavowals.

Given the presence of such examples, we cannot capture what is common to all disavowals as follows: the sentence in the second bit says that the speaker holds such-and-such attitude, which is not the same as the attitude that they have expressed through the first bit. The sentence 'probably $\neg p$' from (15), for example, does not say that the speaker is more confident that $\neg p$ than she is that *p*, or something along these lines. Neither does the sentence 'surely *p*' say that the speaker is sure that *p* in (17).

This makes the task of fleshing out a general explanation of the twisted-ness of disavowals all the more difficult. Difficult, but not impossible.

7 Why direct disavowals sound so twisted

Even though the speaker who utters (15) does not thereby *say that* they are confident that $\neg p$, they do thereby express high confidence that $\neg p$.

And, even though the speaker who utters (17) does not thereby *say that* they are sure that *p*, they do thereby express certainty that *p*. So expression is the more general relation to capture how the second bit of a disavowal (of any kind) relates to its first bit. Let us see how that works for the case of the direct disavowals (10)–(12), as well as the remaining indirect disavowals (13), (14), (16) and (18).

As remarked in §2, the expression of a mental state is a relatively cheap phenomenon. For there are many different speech acts, involving a variety of verbal forms, through which a speaker can express some mental state (which minimally involves, again, their signaling that they are in the relevant mental state to potential hearers—their utterance serves as *evidence* that they are in that state). In particular, the utterance of 'I don't believe that *p*' is not only a means of *saying that* the speaker doesn't believe that *p* (because that is the very bit of information that the target declarative sentence conveys)—but also a means of expressing the state of not believing that *p*.[12] Similarly, 'I am in doubt about whether *p*' is a means of expressing doubt about whether *p*, etc. The difference between 'surely *p*' and 'I am sure that *p*' is that the speaker *says that* they are sure that *p* by uttering the latter, not by uttering the former. But both of them express the speaker's state of being sure that *p*.

Intensional attitude verbs have this special feature, namely, that when a speaker uses the singular first-person pronoun to relate themselves to a proposition or question through that verb, they end up not only *saying that* they have/do not have such-and-such intensional attitude, but also *expressing* that very attitude/its absence to their hearers. 'I intend to be there' is an expression of one's intension to be there, 'I do not intend to be there' an expression of one's lack of intention to be there—even though both of them also say things *about* the speaker's mental state (in contrast to 'You bet!', for example).

In uttering an disavowal, then, the speaker (i) expresses one kind of attitude or mental condition through the first bit of the utterance, and (ii) they express a different kind of attitude or mental condition through the second bit, sometimes using a sentence that also *says that* the speaker has that attitude or satisfies that condition, other times using a sentence that features adverbs such as 'probably' and 'surely'—and one that does

[12] On this point, see also Williams (1998), Marek (2011) and Freitag and Yolcu (2021).

not say that the speaker has that attitude or satisfies that condition.

Now notice that, given the general characterization in (i) and (ii), we cannot expect to account for the twisted-ness of disavowals in the same manner that we have accounted for the twisted-ness of disapprovals above (§4). That is, we cannot explain the twisted-ness of a disavowal starting as follows: assuming sincerity on the part of the speaker, either what is said in the second bit is false, or blah-blah-blah (something about the first bit). For now we cannot simply assume that there is something that is *said* in the second bit of disavowals to begin with, over and above the fact that the speaker expresses some aspect of their mental life through it.

Rather, the relevant explanation should go as follows. Assume that a speaker utters a disavowal. According to our characterization from above, then, the speaker has thereby expressed two different intensional attitudes or conditions in the course of their utterance. The problematic character of disavowals stems from the fact that it is *either* impossible for the utterance to be sincere, in that it is impossible for all of those attitudes or conditions to simultaneously obtain *or*, even if it is possible for them to simultaneously obtain, it is still impossible for all of the attitudes thereby expressed to simultaneously abide by their respective norms or to simultaneously live up to their ideals. In the latter case, again, at least one of the relevant attitudes is guaranteed to be normatively/axiologically subpar, assuming that the other one isn't. Let us examine some concrete examples of this kind of explanation now.

Suppose again the speaker utters (10):

(10 *p*, but I do not believe that *p*.

In uttering the first bit, the speaker expresses an attitude of believing that *p*. In uttering the second bit, the speaker expresses a mental state of non-belief that *p*, signaling that she does *not* believe that *p*. The reason why a speaker's utterance of (1) sounds so twisted is that the speaker cannot satisfy both of these conditions at the same time: they cannot believe that *p* and not believe that *p* at the same time (at least not relative to the same 'fragment'—see Fn. 7). That means, in effect, that it is impossible for the speaker to sincerely utter (10). For, in order for them to sincerely utter (10), they would have to believe that *p* (so as to sincerely utter the first bit), and they would also have to not believe that *p* (so as to sincerely utter the second bit).

A similar explanation holds for the twisted-ness of (11) and (12), since it is impossible for one to be in doubt about whether *p* (a condition expressed by a speaker's sincere utterance of the interrogative '*p?*') and at the same time not be in doubt about whether *p* (a condition expressed by a speaker's sincere utterance of 'I am not in doubt about whether *p*'), and it is impossible for one to intend to be somewhere (a condition expressed by a speaker's sincere utterance of 'I will be there!') and at the same time not to intend to be there (a condition expressed by a speaker's sincere utterance 'I do not intend to be there').

So the twisted-ness of *direct* disavowals stems from the impossibility that the speaker is and is not in a certain mental state at the same time. In the case of *indirect* disavowals, however, the story is a bit different, at least assuming that it is *not* impossible for one to have mutually contradictory beliefs, or to be in doubt about whether *p* while at the same time believing that *p*, or to intend to bring about mutually incompatible scenarios. I turn to that now.

8 Why indirect disavowals sound so twisted

Even assuming that the attitudes that a speaker expresses through an indirect disavowal are compossible, the twisted-ness of their utterance still stems from a certain kind of impossibility, namely, the impossibility that all of those attitudes simultaneously abide by their norms or live up to their ideals. One of the attitudes is guaranteed to be normatively/axiologically subpar, assuming that the other one isn't.

Suppose, for example, that the speaker utters (14):

(14) *p*, but I am in doubt about whether *p*.

In uttering the first bit, the speaker expresses an attitude of believing that *p*. In uttering the second bit, the speaker expresses an attitude of being in doubt about whether *p*. The reason why an utterance of (14) sounds so twisted is that, even if it is possible for the speaker to believe that *p* and be in doubt about whether *p* at the same time, these two attitudes cannot both be at their epistemic bests or abide by their norms at the same time. For one is rationally required not to believe that *p* and be in doubt about whether *p*.[13] And, if one knows that *p*, then one is not justified in being in doubt about whether *p*.

That very same explanation is the one that explains why an utterance of (16) sounds so twisted, again:

(16) *p?* Though I believe that *p*.

For, in uttering the first bit, the speaker expresses an attitude of doubt about whether *p* and, in uttering the second bit, they express an attitude of belief that *p*.

What about indirect disavowals that do not feature intensional attitude verbs, but rather adverbs, in their second bit? Let us consider an utterance of (15), again:

(15) *p*, but probably $\neg p$.

In uttering the first bit, again, the speaker expresses an attitude of believing that *p*. In uttering the second bit, however, the speaker expresses more confidence in $\neg p$ than in *p*, or high credence that $\neg p$ (higher than 0.5, using the unit interval). The reason why an utterance of (15) sounds so twisted is that, even if it is possible for the speaker to believe that *p* and be confident that $\neg p$ at the same time, these two attitudes cannot both be at their epistemic bests or abide by their norms at the same time. For one is rationally required not to believe that *p* and be confident that $\neg p$. And, if one knows that *p*, then one is not justified in being confident that $\neg p$.

Or consider an utterance of (17), again:

(17) *p?* Though surely *p*.

In raising the question in the first bit, the speaker expresses an attitude of being in doubt about whether *p*. In uttering the second bit, however, they express an attitude of being sure *p*. But one of these attitudes is guaranteed to be normatively/axiologically subpar, given that the other one isn't.

[13] This rational requirement is a meant here as a requirement of *coherence*.

So the explanation of the twisted-ness of indirect disavowals—namely, that in uttering them the speaker expresses attitudes that cannot be normatively on the clear or at their bests at the same time—is essentially the same for disavowals that involve intensional attitude verbs and for disavowals that involve adverbs such as ‘probably’ and ‘certainly’.

9 A sum-up of everything

The overarching explanation of the twisted-ness of disavowals of both kinds goes as follows, then. Either they are *twisted ways to speak one's mind*—because they must be insincere, on account of it being impossible for the speaker to hold all of the attitudes that they thereby express—or, even assuming sincerity, they are ways of speaking one's twisted mind—because the attitudes that are thereby expressed are guaranteed to be normatively/axiologically subpar.

Notice that this explanation has the degree of generality that is needed to explain what is wrong about *all* of (10)–(18). Notice, furthermore, that the explanation already contains a diagnosis of what would be wrong with the ‘purely mental’ counterparts of (10)–(18), *if any such mental counterparts could there be.*

The mental counterpart of a disavowal would be a mental state where one holds all the attitudes or satisfies all the mental conditions expressed by the utterance of that disavowal. But in the case of the direct disavowals (10)–(12) and their ilk, as we saw, it is not even in principle possible for a subject to be in a mental state of believing p and not be in a mental state of believing that p, to be in a mental state of doubt about whether p and not be in a mental state of doubt about whether p, etc. So the question concerning the twisted-ness of the purely mental counterparts of direct disavowals doesn't even get off the ground—for *there aren't* such purely mental counterparts to begin with.

Where the purely mental counterpart of a disavowal is even so much as possible, however, the explanation of their twisted-ness is already contained in what was said above: it is impossible for all of the attitudes that make up such counterparts to abide by their norms, or to be at their bests/satisfy their respective ideals at the same time. Where the question about the purely mental counterpart of a disavowal does get off the ground, then, the answer to it can be borrowed directly from the very explanation of the twisted-ness of the *speech act* of uttering that disavowal offered above.

Disapprovals and disavowals are two ways of talking that seem to make manifest some kind of cross-level cognitive dissonance on the part of the speaker, and they constitute a large, comprehensive class of utterances. They are unified by the fact that they feature a step from lower- to higher-order speech: the speaker first says something, and then they say something about their standing with respect to what they said before. I have tried to explain why any member of that large class sounds so twisted in a way that works for both disapprovals and disavowals: they are either twisted ways of speaking our minds (because insincere or inaccurate about our minds) or they are ways to speak our twisted minds (because they express mental states that are guaranteed to fall short of their normative standards or ideals).

Whether this attempted explanation will withstand further critical scrutiny, however, is left for future investigation.

Acknowledgments. I would like to thank Julia Zakkou, Sven Bernecker, Til Eyink, Thomas Grundmann, Kai Wehmeier, Michael Bloome-Tillman and Paul Silva for their feedback on how several examples from (1)–(18) strike them. I am also very thankful to Nadja-Mira Yolcu for her helpful comments on a previous version of this paper.

References

Bar-On, Dorit (2015). 'Expression: Acts, products, and meaning', in S. Gross, N. Tebben and M. Williams (eds.) *Meaning without representation: Essays on truth, expression, normativity*, and naturalism, Oxford: Oxford University Press, pp. 180–209.

Borgoni, Cristina, Dirk Kindermann and Andrea Onofri (2021). *The Fragmented Mind*, Oxford: Oxford University Press.

de Almeida, Claudio (2001). "What Moore's Paradox is About", *Philosophy and Phenomenological Research* 62(1): 33–58.

Elliot, Andrew J. and Patricia G. Devine (1994). 'On the motivational nature of cognitive dissonance: Dissonance as psychological discomfort', *Journal of Personality and Social Psychology* 67(3): 382–394.

Feldman, Richard (2000). 'The ethics of belief', *Philosophy and Phenomenological Research* 60(3): 667–695.

Festinger, Leon (1962). 'Cognitive Dissonance', *Scientific American* 207(4): 93–106.

Freitag, Wolfgang and Nadja-Mira Yolcu (2021). 'An expressivist solution to Moorean paradoxes', *Synthese* 199(1-2): 5001–5024.

Gibbard, Alan (1990). *Wise Choices, Apt Feelings: A Theory of Normative Judgment*, Cambridge, MA: Harvard University Press.

Green, Mitchell and John N. Williams (Eds.) (2007). *Moore's paradox: New essays on belief, rationality and the first person*, Oxford: Clarendon Press.

Lasonen-Aarnio, Maria (2020). "Enkrasia or evidentialism? Learning to love mismatch", *Philosophical Studies* 177(3): 597–632.

Lewis, David (1982), 'Logic for Equivocators', *Noûs* 16(3): 431–441.

Littlejohn, Clayton (2013). 'The Russellian retreat', *Proceedings of the Aristotelian Society* 113(3), 293–320.

Marek, J. C. (2011). 'Expressing and describing experiences: A case of showing versus saying', *Acta Analytica* 26(1): 53–61.

McHugh, Conor and Way, Jonathan (2018) 'What is Good Reasoning', *Philosophy and Phenomenological Research* 96(1): 153–174.

Moore, G. E. (1993). 'Moore's Paradox', in *G. E. Moore: Selected Writings*, ed. by T. Baldwin. London and New York: Routledge.

Rosa, Luis (2022). 'Coherence and Knowability', *The Philosophical Quarterly* 72(4): 960–978.

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

Shah, Nishi (2008). 'How Action Governs Intention', *Philosophers' Imprint* 8: 1–19.

Smithies, Declan (2012). 'The normative role of knowledge', *Noûs* 46(2), 265–288.

Stalnaker, Robert (1984). *Inquiry*, Cambridge MA: MIT Press.

Whitcomb, Dennis (2017). 'One kind of asking', *The Philosophical Quarterly* 67(266): 148–168.

Williams, J. N. (1998). 'Wittgensteinian accounts of Moorean absurdity', *Philosophical Studies* 92(3): 283–306.

Williams, John N. (2015). "Moore's Paradox in Thought", *Philosophy*

Compass 10(1): 24–37.

Williamson, Timothy (2000). *Knowledge and its Limits*, Oxford: Oxford University Press.

Worship, Alex (2018). “The Conflict of Evidence and Coherence”, *Philosophy and Phenomenological Research* 96(1): 3–44.