

BELIEFS OVER AVOWALS: SETTING UP THE DISCOURSE ON SELF-KNOWLEDGE

LUKAS SCHWENGERER

Lukas.Schwengerer@uni-due.de

ABSTRACT

Wright (1998) and Bar-On (2004) put pressure on the idea that self-knowledge as an explanandum should be identified with privileged belief formation. They argue that setting up the discourse on the level of belief and belief formation rules out promising approaches to explain self-knowledge. Hence, they propose that we should characterize self-knowledge on the level of linguistic practice instead. I argue against them that self-knowledge cannot be fully characterized by features of our linguistic practice. I propose that in some circumstances – disagreements about one’s mental states – self-knowledge plays a role, but this role cannot be described in virtue of features of our linguistic practice. I consider three objections to the argument and conclude that we should not conceive self-knowledge solely in terms of linguistic practice.

INTRODUCTION

Wright (1998, 2015) and Bar-On (2004) put pressure on the idea that self-knowledge as an explanandum should be identified with privileged belief formation. They argue that setting up the discourse on the level of belief and belief formation rules out promising approaches to explain self-knowledge. Hence, they propose that we should characterize self-knowledge on the level of linguistic practice instead. I argue against them that self-knowledge cannot be fully characterized by features of our linguistic practice. I propose that in some circumstances – specific kinds of disagreement – self-knowledge plays a role, but this role cannot be described in virtue of features of our linguistic practice.

In section 1 I introduce the choice of setting up the discourse of self-knowledge on either the level of linguistic practice or belief. I thereby explain Wright’s and Bar-On’s setup and their arguments in favor of conceiving the explanandum on the level of our linguistic practice. I then present my argument against them in section 2. I argue that their position lacks the tools to describe the role of self-knowledge in determining what one ought to do if someone disagrees with one’s self-ascription of a mental state. In section 3, I consider three unsuccessful attempts to respond to my argument. I finally conclude that our conception of self-knowledge has to be set up on the level of belief and belief formation to fully capture the phenomenon.

1. SELF-KNOWLEDGE: AVOWAL OR BELIEF?

Generally, I am in the best possible position to say what mental state I am in. If I sincerely say ‘I feel cold,’ or ‘I want a biscuit,’ then rarely anyone ever doubts me. This truism is common

ground in the self-knowledge discourse. Moreover, it is also common ground that there are some differences in our linguistic practice between statements such as ‘I feel cold’ and ‘there is a tree behind the corner.’ It seems perfectly fine to ask ‘How do you know?’ in response to the tree statement, but it is comparatively odd to ask the same in response to my claim about feeling cold. It seems inappropriate to question the mental state self-ascription. It seems inappropriate because the natural answer to the question is a reiteration of the initial claim. I know that I am feeling cold, because I *do feel cold* – I just know!

Examples like these are a common way to introduce the idea that self-knowledge is *special* and *privileged*. It is special because the way I can avow my own mental states differs from attributions of mental states to other people. Moreover, it is privileged because I am the authority regarding my mental states. However, this way of introducing self-knowledge is less straightforward than one might expect. Here these distinctive features of self-knowledge are made salient by speech acts of authoritative, psychological self-ascription. Call these self-ascriptions ‘avowals.’ The question is whether features of self-knowledge should be identified wholly with our linguistic practice of avowing, or whether we require something beyond speech acts. We want to explain self-knowledge, but what exactly is it that we want to explain? Wright formulates this choice:

So the would-be theorist of self-knowledge confronts a fork. What comes first here in the order of explanation: the linguistic practice, or the thoughts of the thinkers manifested in that practice? The problem of self-knowledge will look very different depending on how we choose. (Wright 2015: 52)

As Wright indicates it is a question about the starting point for any inquiry into self-knowledge. Even if we all agree that there is something special and privileged about self-knowledge, it is unclear where these features are located. Moreover, this decision of locating self-knowledge determines how the features can be explained. If we take self-knowledge to be an instance of special and privileged belief we have to explain what makes it special and privileged in terms of properties of belief and belief-formation. If we locate it at the level of linguistic practice, as an instance of avowals as a special kind of speech act, we have to explain the specific rules governing the speech act. In this case we need to explain what makes the speech act special. Hence, Wright rightfully emphasizes that this choice of a starting point needs to be properly addressed. We are confronted with a choice between a *linguistic view*, and a *doxastic view*.¹ We can capture the two options with these two principles:

(Linguistic View) The peculiar nature of self-knowledge should be described exclusively by features of linguistic practice, syntax, semantics, and pragmatics.

(Doxastic View) The peculiar nature of self-knowledge should be described exclusively by features of beliefs and belief formation.

(Linguistic View) expresses the language-first path of the fork, whereas (Doxastic View) captures the thought-first path. These views are defined in terms of a description to keep the characterization of the explanandum distinct from possible explanations for

¹ This terminology is mine, not Wright’s or Bar-On’s.

the respective explananda. They are views about the proper choice of an explanandum first, and only derivatively tell us anything about explanations. Clearly, the explanandum restricts possible explanations, but only to some extent. For instance, an explanandum on the level of language can still be compatible with an explanation on the level of belief. On the other hand, an explanandum on the level of belief is not compatible with an explanation on the level of language. We will come back to this point when discussing the proposed advantages of the linguistic view later, because a setup based on the linguistic view is supposed to be compatible with more types of explanations than a setup based on the doxastic view.

Both the linguistic view and the doxastic view presuppose that their suggestion can actually be followed through. We therefore have the corresponding presuppositions in place:

(Linguistic Presupposition) The peculiar nature of self-knowledge can be described exclusively by features of linguistic practice, syntax, semantics, and pragmatics.

(Doxastic Presupposition) The peculiar nature of self-knowledge can be described exclusively by features of beliefs and belief formation.

Moreover, both views aim to capture our folk notion of self-knowledge. According to the linguistic view the way in which self-knowledge is special and privileged is to be spelled out on the level of speech acts without missing out on any feature we would pre-theoretically attribute to self-knowledge. To do so a proponent to the linguistic view might identify the explanandum with a set of features of avowals. The doxastic view on the other hand characterizes the intuitive features of self-knowledge as results of a special and privileged belief-formation.

A good example of a version of the doxastic view can be found in Alex Byrne's characterization of self-knowledge²:

First, knowledge of one's mental states is *privileged* in comparison to knowledge of others' minds. Roughly: beliefs about one's mental states acquired through the usual route are more likely to amount to knowledge than beliefs about others' mental states (and, more generally, beliefs about one's environment). ... Second, knowledge of one's mental states is *peculiar* in comparison to one's knowledge of others' minds. One has a special method or way of knowing that one believes that the cat is indoors, that one sees the cat, that one intends to put the cat out, and so on, which one cannot use to discover that someone else is in the same mental state. (Byrne 2005: 80–1)

The features of self-knowledge are described in terms of the high likelihood of beliefs about one's mental states to amount to knowledge, and with reference to the specific belief-forming process that generates beliefs about one's mental states.

My main aim in this paper is to provide an argument against the linguistic view. Therefore, this example will be sufficient to illustrate the doxastic view. I now sketch

2 A reviewer rightly points out that this is merely Byrne's characterization of the 'problem of self-knowledge' as explanandum and need not capture self-knowledge exclusively. I agree with this remark. Nevertheless, for my purpose Byrne's setup is sufficient to illustrate how doxastic features can be used as an explanandum for the peculiar nature of self-knowledge.

two different ways of developing the linguistic view to provide a better idea of what this alternative setup for self-knowledge might look like. I start by discussing Wright's position, and then continue with Bar-On's alternative.

Wright (1998, 2001, 2015) proposes *immediacy, authority, and salience* as the explanandum for self-knowledge.³ These are features of avowals defined with reference to appropriate speech acts and responses. The exact nature of them differs depending on the type of avowal in question. Wright distinguishes phenomenal and attitudinal avowals. Both categories are not explicitly defined, but rather explained by a selection of examples. Phenomenal avowals are illustrated with examples like 'I have a headache,' 'My feet are sore,' or 'I'm tired' (Wright 1998: 14).

Phenomenal avowals⁴ show three key features (Wright 1998: 14–15):

- *Immediacy*, which captures the observation that demanding reasons or evidence from a person that voices an avowal seems inappropriate. The question 'How can you tell?' is always inappropriate as a response to a phenomenal avowal.
- *Strong authority*, which means that whenever someone sincerely claims that she is in mental state *x*, and understands what this claim means, this guarantees that the claim is true. Any doubt about such a claim has to be a doubt about sincerity or understanding.
- *Salience*, defined by the absurdity of avowing uncertainty about one's own mental states. If one is asked 'Do you have a headache?' it seems absurd to answer 'I don't know.'

The second category of speech acts of self-ascriptions are attitudinal avowals. These are avowals of content-bearing states, such as 'I believe that term ends on the 27th,' 'I hope that noise stops soon' (Wright 1998: 15). Attitudinal avowals differ to phenomenal avowals insofar as they are only weakly authoritative. They provide empirically assumptionless justification for the corresponding third-person claims. However, one can (even though one rarely does) doubt individual attitudinal avowals without doubting sincerity or understanding. Nevertheless, one cannot doubt that a person correctly avows attitudes in general. Wright labels speech acts with weak authority as *inalienable*. That is, one cannot be always unreliable about one's own mental states. Generally, when someone avows that she believes *p*, we trust that she is right. We have a presumption that subjects tell the truth, even though there is no guarantee when they avow attitudes. Finally, attitudinal avowals also show a slight difference with regard to salience. If one is asked 'Do you believe that *p*?' it seems absurd to answer 'I don't know.' However, for attitudinal avowals it might not be absurd to answer that one suspends judgment. It would only be absurd if you did not know whether you believe that *p*, not believe that *p*, or are withholding judgment.

Bar-On's (2004) formulation of the explanandum 'self-knowledge' looks similar to Wright's conception. She also formulates it in terms of features of avowals which are indicated in the desiderata her account aims to explain (she states 8, I only mention the 4 that capture epistemic asymmetry):

DI. The account should explain what renders avowals protected from ordinary epistemic assessments (including requests for reasons, challenges to their truth, simple correction, etc.).

³ Wright (1998, 2001) uses *groundlessness* instead of immediacy, and *transparency* instead of salience.

⁴ For criticism on Wright's characterization of phenomenal avowals see Snowdon (2012).

D2. It should explain why avowals' security is unparalleled: why there are asymmetries in security between avowals and all other empirical ascriptions, including (truth-conditionally equivalent) third-person ascriptions and non-mental first-person ascriptions. In particular, it should explain why avowals are so strongly presumed to be true.

D3. It should explain the non-negotiable character of the security – the fact that it is 'non-transferable' and 'inalienable.'

D4. It should apply to both intentional and non-intentional avowals alike, and allow us to separate avowals from other ascriptions in terms of their security. (Bar-On 2004: 20)

Bar-On motivates these features by pointing to ordinary examples. In everyday cases no one questions my avowals, and if they do it usually seems inappropriate. On the other hand, my assertions about external objects are comparatively often questioned.

There is no feature parallel to Wright's salience idea in Bar-On. In her characterization it seems perfectly fine to respond with 'I don't know' to a question of whether one is in pain, or whether one believes something. This might strike one as odd, partially because of phenomenological observations, and partially because of the emphasis on salience in discussions in the epistemology literature (e.g. debates on Williamson's (2000) anti-luminosity argument). Moreover, it might lead to problems accounting for negative avowals – avowals of what mental state one is not in (Brueckner 2011). For instance, we seem to be able to make especially secure avowals of not being in pain, or not believing that p. It is at least not obvious how these avowals fit into Bar-On's framework.

Bar-On's discussion of self-knowledge comes with a further complication. While she does subscribe to the linguistic view for the most part,⁵ she does not fully commit to this position. Bar-On distinguishes three different questions related to self-knowledge.

- (i) What accounts for the *unparalleled security of avowals*? Why is it that avowals, understood as true or false ascriptions of contingent states to an individual, are so rarely questioned or corrected, are generally so resistant to ordinary epistemic assessments, and are so strongly presumed to be true?
- (ii) Do avowals serve to articulate *privileged self-knowledge*? If so, what qualifies avowals as articles of knowledge at all, and what is the source of the privileged status of this knowledge?
- (iii) Avowals aside, what allows us to possess privileged self-knowledge? That is, how is it that subjects like us are able to have privileged, non-evidential knowledge of their present states of mind, regardless of whether they avow being in the relevant states or not? (Bar-On 2004: 11–12)

As a starting point (iii) is discarded because it assumes privileged self-knowledge and thereby also denies deflationary answers to (i) and (ii). Deflationary here means that any privilege for self-knowledge is understood exclusively on the level of linguistic practice, and any doxastic privilege is rejected. The methodologically interesting decision is

5 Moreover, Wright explicitly names Bar-On as a paradigmatic case of the language first path of his fork (Wright 2015: 52).

to pick (i) over (ii). Bar-On believes that the interest in the phenomenon of self-knowledge arises from the asymmetry to knowledge of others, and this asymmetry is directly taken from the special nature of avowals in conversation and thought. We need to start at the avowals, because they determine our perspective of self-knowledge. Starting at avowals with question (i) is the move that fits the linguistic view. The question (i) is formulated in terms of linguistic practice, and thereby the explanandum is the set of language based features D₁–D₄. However, she does not commit to answers to (ii) and (iii). She rather provides different possible answers, including some deflationary answers that reject privileged self-knowledge on a doxastic level. Thereby Bar-On (2004: Ch. 9) provides some combination of these answers to (ii) and (iii) that fit with the linguistic view, and others that do not.⁶ Some answers that do not fit with the linguistic view lead towards a hybrid view between linguistic and doxastic. For instance, in one option the notion of belief is weakened such that an avowal itself is enough to constitute a belief (Bar-On 2004: 365). In this option the clear distinction between linguistic and doxastic is lost. Because she discusses various different possible views it is difficult to evaluate Bar-On's overall picture of self-knowledge. Nevertheless, the special security of self-knowledge is fully understood in terms of the linguistic view. Moreover, it is clear that Bar-On does not want to rule out a deflationary answer to question (ii) in her initial set-up. Cautiously formulated we can say that there is at least one Bar-On *inspired* position that is fully in the spirit of the linguistic view. This is a view that understands the explanandum for (i) with Bar-On according to the linguistic view, and provides deflationary answers to (ii) and (iii).

With Wright's and Bar-On's variations of the linguistic view on the table,⁷ we might ask which formulation of self-knowledge under the linguistic view is best. However, I will bracket this question and not take a stance on which formulation is preferable. Instead I focus on the more general principle (Linguistic View). Why should we choose the linguistic view instead of the doxastic one? Both Wright and Bar-On bring a similar line of argument to the table: We should prefer the linguistic view over the doxastic view, because the latter skews the discourse in a way that rules out some promising explanations of self-knowledge.

Bar-On (2004) argues that starting with questions about beliefs and knowledge brings about a dangerous temptation to only see one possible account of self-knowledge: some kind of especially secure method of making judgments about our present states of minds. If we start looking at doxastic states we will be blind to non-epistemic explanations of self-knowledge. Moreover, we will be tempted to overly assimilate avowals to assertions, which in turn might bring about problems to adequately explain what is special about self-knowledge. If we adopt the doxastic view we are forced to an epistemic approach. With the linguistic view on the other hand "we begin with a relatively neutral set of observations about the status of avowals, and try to understand that status We can then take up questions of self-knowledge with a more open mind" (Bar-On 2004: 13).

Wright (2015) has a similar aim in mind. Responding to Snowdon (2012) he provides the following rationale for the linguistic view:

6 See also Bar-On (2011) where she reaffirms that she does not argue for any settled view on privileged self-knowledge.

7 And these are not the only options. Others include Ryle (1949 [1984]), and Finkelstein (2003).

To set the record straight, then: I did not mean, in the Whitehead lectures, to side with Wittgenstein's view of these matters. But I did want to set things up in a way that allows his view to be heard. (Wright 2015: 55)

Here Wittgenstein serves as a placeholder for non-epistemic accounts in general. Wright's charge against the doxastic view is that any account that is not purely epistemic cannot enter the discourse if we set it up according to the doxastic view. Take for instance Wright's 'default view' as one such option. This is the idea that the authority of avowals is a constitutive principle that is not the result of any epistemic relation. He characterizes this view as follows:

[T]he authority standardly granted to a subject's own beliefs, or expressed avowals, about his intentional states is a *constitutive principle*: something which is not a consequence of the nature of those states, and an associated epistemologically privileged relation in which the subject stands to them, but enters primitively into the conditions of identification of what a subject believes, hopes and intends. (Wright 1989: 142)

Wright is correct; this is not an available option if one starts with the doxastic view. The doxastic view presupposes that the peculiar nature of self-knowledge can be described exclusively on the level of belief and belief-formation. However, the default view explicitly rejects this option by stating that authority is something that avowals have by default, just by being avowals. There is no further fact that brings about the authority. Authority is a feature of the logical grammar of our speech. This proposal that the authority of avowals is something primitive to the speech act has no place in the doxastic view, because it posits that the peculiar nature of self-knowledge can only be explained as a linguistic feature of avowals. Only if the starting point is one of linguistic practice can this feature be primitive. Regardless of whether Wright's approach is successful, leaving the option on the table is *prima facie* a virtue of the linguistic view. Furthermore, this line of argument is strengthened by Wright's claim that the linguistic view provides more options, while supposedly not rejecting any options that are permissible under the doxastic view. The linguistic view does not entail that

(Language Only) the peculiar nature of self-knowledge can *only* be explained by features of linguistic practice, syntax, semantics, and pragmatics.

The linguistic view merely claims that self-knowledge can be described in terms of linguistic practice, not that this description is irreducible. Perhaps we can still explain the linguistic practice in terms of privileged belief formation. Hence, Wright states that if we set up the explanandum in terms of avowals, an "explanation in terms of cognitive advantage is by no means thereby ruled out . . ." (Wright 2015: 54). However, with the linguistic view other explanations, such as Wright's default view or neo-expressivist⁸ solutions, become available. A similar line of thought can be found in Bar-On, who holds that our account

8 Expressivists treat avowals similar to natural expressions like grimaces or crying: they function to express a mental state of a person rather than reporting the mental state. A brute form of expressivism denies that avowals are anything besides expressions. Crucially, they are not assertions at all (Wright 1998: 34). Neo-expressivists on the other hand propose that avowals can have this expressive function, while still being truth-evaluable. Cf. Finkelstein (2003), Bar-On (2004), Bar-On and Sias (2013).

of the distinctive security of avowals should ideally leave a non-deflationary view of self-knowledge open. That is, that a description of the peculiar features of self-knowledge in terms of our linguistic practice should be compatible with an account on the level of belief (Bar-On 2004: 20). Based on her neo-expressivist account of avowals she even provides different ways in which a non-epistemic explanation of the features of our linguistic practice could be connected to privileged beliefs (Bar-On 2004: Ch. 9). However, just like Wright, Bar-On thinks the best starting point are avowals and their features, so that non-epistemic explanations of the peculiar nature of self-knowledge are available in the first place. What we should explain is self-knowledge as a phenomenon on the level of our linguistic practice – especially secure avowals. If it turns out that this explanation fits with privileged belief states this is a welcome result, but the explanandum itself does not necessarily require anything on the level of belief. The peculiar nature of self-knowledge has to be understood in a way that does not rule out a negative answer to the question whether one has privileged beliefs about one's own mental states. Bar-On herself does not give this negative answer (2004: 24), but her description of the especially secure nature of avowals is compatible with it.

The commitments of the linguistic view are supposedly small. Nevertheless, proponents of the linguistic view are committed to (Linguistic Presupposition): the possibility of a description of the peculiar nature of self-knowledge by features of linguistic practice, syntax, semantics, and pragmatics. Furthermore, from this presupposition follows that any explanatory role that self-knowledge plays can be described by reference to features of linguistic practice, syntax, semantics, and pragmatics. Call this principle (Linguistic Features):

(Linguistic Features) Any explanatory role that self-knowledge plays can be described by reference to features of linguistic practice, syntax, semantics, and pragmatics.

(Linguistic Features) is satisfied by both Wright's and Bar-On's view. They agree that self-knowledge talk can in principle be translated to talk about features of our linguistic practice. That I have privileged knowledge about myself can be translated to my avowals being protected from ordinary assessment, or the avowals having authority. If I say that I am in pain, my avowal will be accepted in normal circumstances.

2. THE ARGUMENT AGAINST THE LINGUISTIC VIEW

My aim is to challenge the linguistic view by targeting (Linguistic Features). The strategy is to find a feature of self-knowledge that cannot be spelled out in terms of linguistic practice. This undermines the linguistic setup by putting pressure on the idea that the linguistic view fully captures our folk notion of self-knowledge. There might still be a hybrid view that accepts self-knowledge as having features on multiple levels, but setting up the problem solely on the level of linguistic practice will be hopeless. If the argument is successful it thereby threatens Wright's position and the neo-expressivist project by undermining their starting point. In a nutshell the argument works as follows:

P1: The features of self-knowledge play a role in philosophical problem X, so they should tell us something about problem X.

P₂: No feature of our linguistic practice is going to tell us anything about philosophical problem X.

C: Self-knowledge cannot be captured wholly in features of our linguistic practice.

Given (Linguistic Features) any explanatory role that self-knowledge plays can be described in terms of features of our linguistic practice, syntax, semantics, and pragmatics. Now suppose a case of a philosophical problem in which intuitively self-knowledge plays an explanatory role. Based on (Linguistic Features) we should be able to re-describe the role of self-knowledge in this case in terms of features of the linguistic practice. However, if there is no adequate description that substitutes the folk notion of self-knowledge with features of our linguistic practice, then it looks like self-knowledge cannot be captured exclusively by features of our linguistic practice. Hence (Linguistic Features) must be false. Moreover, because (Linguistic Features) is an entailment from (Linguistic Presupposition) we can conclude that (Linguistic Presupposition) is false, and the linguistic view collapses.

The problem in question is the case of a *mental state disagreement*⁹: A subject S believes (by non-interpretational means¹⁰) that S is in mental state M. An interlocutor L claims that S is not in M, but rather in a different state M*. We then ask, whether S ought to change her confidence in her belief in the face of disagreement. To illustrate mental state disagreements consider the following case:

(Friend) Suppose I am part of a university admission committee. I know that I am supposed to be fair towards all applicants. I explicitly state that I treat every submission the same, regardless of the ethnicity of an applicant. “I believe ethnicity makes no difference in the quality of a candidate,” I say. However, my friend disagrees. She knows me well and she also knows all my 25 past decisions on the committee. We disagree about my mental state. Being confronted with this disagreement, what am I to do?

In this case it seems plausible to concede that I might be biased.¹¹ Perhaps I actually believe that people of my ethnicity are better applicants, but I lack awareness of this belief. Given that my friend knows all my past decisions, it seems that I should accept her testimony and lower the confidence in my belief.

We can further characterize mental state disagreements. First, I can be wrong in assessing my mental states, and my interlocutor can be right. I do not argue for this here, but merely point out that it is accepted by proponents of the linguistic view (cf. Wright 1998, 2001, 2015; Finkelstein 2003; Bar-On 2004).

Second, interlocutors in mental state disagreements are not epistemically equal. My friend and I form our beliefs on a different basis. She observes my behavior and infers

9 I do not claim that is the only problem that can be used for this line of argument.

10 This clause is supposed to rule out cases in which I believe that I am in a mental state based on observing and interpreting my behavior. Some philosophers argue that there is no non-interpretational self-ascription of attitudes. They propose that self-ascriptions are based on interpretation, even though we might be in a better position to observe ourselves than others. Cf. Carruthers (2011), Cassam (2014).

11 This presupposes that actions are a guide to belief. I take this to be a plausible assumption, although not completely uncontroversial.

my belief, I introspect. We are also aware that we form our beliefs differently and are not epistemic peers. Moreover, interlocutors in mental state disagreements do not only happen to be epistemically unequal, they are in principle unequal. They cannot get on the same epistemic level by disclosing their evidence. This inability is one-sided. My friend has no problem to disclose her evidence. She can state what behavior she observed. However, I am unable to disclose my evidence. Whatever the basis for my belief is (if there is one), I cannot access it. My only option appears to be stating that ‘I just know.’ Full disclosure is ruled out. Hence, there is no way to get on an equal epistemic footing.

One might suggest that this is not a problem. Even though full disclosure is ruled out in the sense that I cannot share my evidence, it should be enough that my interlocutor can cite her evidence. I can then adjust my belief on the total evidence, mine and hers. However, it is unclear how exactly this is supposed to work. I cannot weight her evidence against mine, because that would also require access to my evidence for my mental state.

Third, there is no universal ideal response to all mental state disagreements. In (Friend) I ought to change my confidence in my second-order belief. However, consider a slightly different case:

(Passer-by) Suppose I am part of a university admission committee. I know that I am supposed to be fair towards all applicants. I explicitly state that I treat every submission the same, regardless of the ethnicity of an applicant. “I believe ethnicity makes no difference in the quality of a candidate,” I say. A passer-by disagrees. She does not know me well, and she only knows a single decision of mine on the committee. Being confronted with this disagreement, what am I to do?

It seems out of the question to revise my belief here. My interlocutor barely knows me, and she has little evidence for her judgment. I ought to stick to my belief and confidence level. Because some cases require one to lower one’s confidence, while other cases rationally require one to hold onto one’s confidence level there clearly is no universal response to mental state disagreements. The difference between (Friend) and (Passer-by) seems to be the epistemic standing of my interlocutor. The more evidence my interlocutor can cite, the more rational it seems to decrease confidence in my belief. This is not surprising as it is a feature of disagreements in general. The more justified I take a disagreeing interlocutor to be, the more rational it is to adjust my belief based on the disagreement. Furthermore, I have to believe that my interlocutor surpasses some threshold of justification before the disagreement rationally requires me to decrease confidence in my belief at all. As long as I do not take the interlocutor to pass this mark, the disagreement is ineffective.

In mental state disagreements the threshold for rationally required change in confidence seems to be higher than in ordinary disagreements. This can be illustrated by considering two parallel cases, one involving disagreement about a third person’s mental state, and one involving disagreement about my mental state.

(a) Suppose I am at a party. Kate, a friend of mine, and John, my long-time colleague, are also present, but they are in a different room at time₁. Later at time₂ I talk to Kate about 80s music. I sincerely say that John believes 80s music is terrible. Kate, who observed John picking out music at the party at time₁, disagrees. She thinks John

actually believes that 80s music is good, and it shows in his behavior selecting typical music from the 80s at the party. However, Kate only met John 3 months ago, and hence does not know John very well. Should I lower the confidence in my belief that John believes 80s music is terrible?

- (b) Suppose I am at a party, choosing some music to play. Kate, a friend of mine, is also present. Later I talk to Kate about 80s music. I sincerely say that I believe 80s music is terrible. Kate, who observed me picking out music at the party earlier, disagrees. She thinks I actually believe that 80s music is good, and it shows in my behavior selecting typical music from the 80s at the party. However, Kate only met me 3 months ago, and hence does not know me very well. Should I lower the confidence in my belief that I believe 80s music is terrible?

I suggest that the answer to (a) is yes, and the answer to (b) is no. In the latter case I know myself better than my friend knows me. I know myself better to a degree that makes it permissible to disregard her disagreement. However, I cannot disregard her testimony about my colleague in case (a), because I do not know my colleague in a similar way. The disagreement about someone else's mental state requires less to be effective than the disagreement about my own mental state. If this is correct, then it seems that the involvement of self-knowledge makes a difference for determining the rational response to a disagreement. This is the crucial step in the argument against the linguistic view. Once you accept that intuitively self-knowledge plays a role in determining the rational response to mental state disagreements the linguistic view is in trouble.

Given that self-knowledge plays this role in mental state disagreements we can search for ways in which self-knowledge might play this role. The obvious choices to look into are justification and confidence. If self-knowledge is thought of as a product of peculiar belief formation that provides especially strong justification, then the influence on a rational response to disagreements is straightforward. I am more justified in my mental state self-ascriptions than I am in attributing mental states to others. Hence, I can stick to my belief in (b), but cannot do so in (a). This is a perfectly fine explanation for advocates of the doxastic view. However, for proponents of the linguistic view the answer cannot be justification or confidence. They are committed to (Linguistic Features), the claim that the role that self-knowledge plays can be described by reference to features of the linguistic practice, and neither justification, nor confidence¹² in beliefs seem to be part of the linguistic practice. The challenge is that they need to find a way in which my rational response is influenced by self-knowledge without relying on any epistemic differences between (a) and (b). This challenge seems impossible to meet. Any characterization of self-knowledge in terms of linguistic practice will be quiet on how I ought to rationally change my beliefs in any situation. After all, rational beliefs are not part of the linguistic practice.

The problem is that the linguistic view defines the features of self-knowledge in terms of our linguistic practice with reference to appropriate or inappropriate speech acts, either avowals, or responses to avowals. However, in a mental state disagreement we already start with an avowal and a response. Whether they are appropriate does not seem to matter at this point. All we are interested in is what to do *after* the initial response by my

12 Bar-On (2004) mentions a high degree of confidence of one's avowal in her introduction. However, this cannot be an epistemic notion of confidence if she wants to hold on to the linguistic view. Hence she cannot use confidence in beliefs to explain the difference between (a) and (b).

interlocutor. All linguistic features are already out of the game once the disagreement enters. Nevertheless, self-knowledge seems to play a role here, because our rational response can differ between mental state disagreements and corresponding ordinary disagreements.

Take a second comparison. This time let us compare a disagreement case involving an interpretation based self-ascription with a disagreement involving genuine self-knowledge.

- (c) Suppose Anna and I are candidates for a job. Anna gets the job, while I have to keep looking for work. In the next days I notice that I act a little hostile towards Anna. My parts in our conversations are short and my tone is rather unfriendly. Looking at my own behavior I conclude that I must be envious. Talking to Anna I apologize and tell her that I'm envious which is why I act so rude. She disagrees, telling me that I'm just frustrated that I have to keep looking for a job, I'm not really envious. Anna knows me well. Moreover, I know that Anna is always blunt and is sincere in her assertion. Should I lower the confidence in my belief that I'm envious?
- (d) Suppose Anna and I are candidates for a job. Anna gets the job, while I have to keep looking for work. In the next days I act a little hostile towards Anna. My parts in our conversations are short and my tone is rather unfriendly. Without noticing this behavior I believe that I'm envious of Anna. Talking to her I apologize and tell her that I'm envious. She disagrees, telling me that I'm just frustrated that I have to keep looking for a job, I'm not really envious. Anna knows me well. Moreover, I know that Anna is always blunt and is sincere in her assertion. Should I lower the confidence in my belief that I'm envious?

I suggest that just like in (a) and (b), the answer to (c) is yes, and the answer to (d) is no. One might suspect that the linguistic view has the tools to deal with this case, because its proponents make a distinction between proper avowals and assertions based on self-interpretation. However, the linguistic view still lacks the tools to tell us why lowering the confidence in my belief would be *rational* in one case, but not in the other. All the linguistic view can tell us here is that in (c) Anna's disagreement is appropriate, but in (d) it is not. Once again, the linguistic view does not tell us anything about our response after Anna actually disagreed, but self-knowledge still plays a role at this point. The linguistic view fails to fully capture our folk notion of self-knowledge.

3. AGAINST THE ARGUMENT

Proponents of the linguistic view cannot respond by simply denying that mental state disagreements exist. They definitely do, and moreover they seem unavoidable if one accepts fallibility for self-ascriptions. However, I think there are at least three different, interesting ways to respond to the argument for the proponents of the linguistic view. First, they may object that I underdescribe the cases and hence we cannot be sure what our intuitions should be here. Second, they may argue that I smuggled the doxastic view into my premises when I set up the problem as a question about a rational response to a mental state disagreement related to my confidence in my belief. Third, they may deny that genuine self-knowledge plays a role in these disagreements.

3.1 *The case is underdescribed*

Friends of the linguistic view can disagree with my suggested answers for cases (a) to (d). Furthermore, they might contest that it is generally unclear what to think about these cases, because they are not sufficiently well described. In (a) and (b) we don't know exactly how well people know each other. John is a long-time colleague, but what exactly does that mean for my knowledge of John's typical behavior? Perhaps I know John so well that I can safely ignore Kate's disagreement. Or perhaps Kate knows neither John nor me well enough for her disagreement to matter. Similarly, in (c) and (d) the cases do not state how well exactly Anna and I know each other, nor what my behavior in the case is exactly.

However, I do not think that any charitable way of filling in more details is going to change the intuitions pumped. All that is required for the argument is that we compare a case in which the belief was intuitively formed on observation, inference and interpretation, with a case that is intuitively an instance of self-knowledge (or at least self-belief). The latter has to be stated such that it is easily recognized as an instance that does not appear to be based on observation, inference and interpretation. As long as we hold this difference in intuitive belief-formations fixed we can add as many details as we want. For instance, I can add that Anna and I have known each other for three years and we meet about once a week to (c) and/or (d). I might further add that we usually have long conversations, especially about topics we care about. Perhaps, we talk about a topic I usually am enthusiastic about, but nevertheless I clearly attempt to end the conversation as quickly as possible. These additions are not changing the intuitions pumped, even in case they introduce differences between (c) and (d). The intuition pumped is still that I seem to require different adjustments to my beliefs in (c) compared to (d). Hence, the charge of underdescription seems to miss the point.¹³

3.2 *Begging the question*

The second objection raises an issue about the setup of my argument. I ask what is rational to do in mental state disagreements. Should I lower the confidence in my belief or should I stay put? However, this already locates the discussion to the level of belief, the proponent of the linguistic view might object. They would describe mental state disagreements differently: A subject *S* avows that *S* is in mental state *M*. An interlocutor *L* claims that *S* is not in *M*, but rather in a different state *M**. We then ask, whether it is appropriate for *S* to avow that *S* is in mental state *M* in the face of disagreement.

The immediate response here is to ask whether this is the right way to describe mental state disagreements. In ordinary disagreements it is fine to ask the question of what one is supposed to believe in the face of disagreement, so why should it not be equally fine to ask the question when the disagreement is about my mental states? Moreover, based on our ordinary linguistic practice there does not seem to be anything wrong with wondering what one ought to believe in a mental state disagreement. So this objection has to be further motivated to get off the ground.

13 Thank you to an anonymous reviewer for pointing out that my response to this point in an earlier version was misdirected, and that a straightforward answer allowing details to be added in any charitable way is available.

Let us suppose it can be sufficiently motivated. Even in this case, we can run a version of my overall argument, because in some disagreements of this kind it seems appropriate for S to still avow that S is in mental state M, while in others it is not. And the difference between these cases has to be accounted for without any epistemic difference, which seems to be challenging.

However, perhaps even this description of mental state disagreements is not acceptable for the linguistic view philosophers. They could argue that the initial disagreement itself was inappropriate by the interlocutor L. Furthermore, the question how one appropriately responds to an inappropriate speech act is misguided. No inappropriate speech acts demands a particular response. One cannot be blamed for any response to an inappropriate speech act. There is no rule in our linguistic practice to govern inappropriate challenges of my avowal. Once my opponent stops playing by the rules I cannot look at the rules for what to do. Just as there is no legitimate chess move as a response to someone stacking the Knight on top of the Rook, there is no proper move in the language game after a mental state disagreement.

The problem with this response is that mental state disagreements do not look like a complete breakdown of the rules of communication. Mental state disagreements can be appropriate, even though they are rare occurrences. Moreover, when someone disagrees about my mental state, there appears to be a right and a wrong response for the particular case. We have not stopped playing our language game. I take this to be an indication that our folk notion of self-knowledge includes the possibility of challenges by others. There is a right response to such challenges, and our theoretical conception of self-knowledge should provide enough tools to fully explain why we ought to respond a certain way. The linguistic view cannot do that, and hence fails to capture our folk notion of self-knowledge.

3.3 *No genuine self-knowledge*

Finally, friends of the linguistic view may deny that genuine self-knowledge plays a role in mental state disagreements. One can make the case that Bar-On (2004) has a response of this kind built into her account. Her conception of the explanandum for self-knowledge involves avowals being “protected from ordinary epistemic assessments (including requests for reasons, challenges to their truth ... etc.)” (p. 20). This gives her the option to rule out mental state disagreements as cases of *extraordinary* epistemic assessments. Furthermore, she provides examples of extraordinary mental state self-ascriptions, including “on the basis of therapy, *consultation with others*, self-interpretation, or cognitive test results” (p. 194, emphasis added). On this basis Bar-On can argue that the question of a rational response to mental state disagreements is not relevant for genuine self-knowledge, because the mere fact that I take the interlocutor’s disagreement seriously indicates that I left the ordinary linguistic practice of avowing, and entered a different language game. One in which my statement looks like an avowal, but is treated as a mere report. In other words, Bar-On can contest whether the impact of self-knowledge in mental state disagreements is actually part of our folk notion of self-knowledge.

This response seems to presuppose a too narrow scope of self-knowledge. It seems arbitrary to posit that certain sincere and largely non-interpretative mental state ascriptions are not avowals, while others are. Bar-On might respond that it is not arbitrary for

two reasons. First, because in mental state disagreements I consider evidence my interlocutor provides. My self-ascription after taking my interlocutor's disagreement seriously appears to be partially inferred from evidence. Hence, it will be different than ordinary avowals which appear to be non-evidential (Bar-On 2004: 2). Second, all avowals show immunity to error through misascription. They do not involve any recognition of a mental state and therefore are protected from epistemic assessment (Bar-On 2004: Ch. 6). My response to the mental state disagreement on the other hand is an attempt to recognize my mental state correctly with the help of my interlocutor's testimony, hence it is not an avowal.

However, both reasons can be challenged. First, even though the interlocutor's testimony constitutes evidence, the mental state ascription can still be largely non-evidential. I am not self-ascribing a belief *only* on the testimony. Plausibly I can have my own, non-evidential judgment, which I then adjust based on the testimony. The result appears neither fully evidential, nor fully non-evidential. The question is whether we should treat it like the fully evidential, or like the fully non-evidential case.¹⁴ I think there is a prima facie reason against the former. After I adjust my belief according to the disagreement I can still be the authority regarding my self-ascriptions. This authority is something that is not present in case I assert solely based on my evidence. There is still an asymmetry between the self-ascription in the post-testimony case and ascriptions of others' mental states (or ascriptions of my own mental states based fully on evidence). If this is correct, then we should treat the self-ascription after considering testimony more akin to genuine avowals.¹⁵

Second, arguing from the immunity to error through misascription gets the order of explanation wrong. The concept of immunity to error through misascription is introduced by Bar-On (2004: Ch. 6) to explain why avowals are protected from ordinary epistemic assessment. Avowals, so Bar-On claims, do not involve any recognition of a mental state. Moreover, you cannot accuse me of making an epistemic mistake, if I did not perform any epistemic action at all, so challenges to my avowal are off the table. However, given that this feature is supposed to explain a property of avowals we cannot use it to pick out which speech acts are avowals. We want to find out whether an explanation in terms of immunity to error through misascription fits the folk notion of self-knowledge, and therefore we should not pick out the extension of this folk concept in virtue of the theoretical concept in question.

Given that these reasons to disregard the responses to mental state disagreements as cases of genuine self-knowledge are not well motivated we are left without any principled way to rule out mental state disagreements as ordinary interactions involving self-knowledge. Hence, we are stuck with appealing to intuitions. And it does not seem intuitive to treat my claim 'I believe that p' differently as soon as someone disagrees and I take the disagreement seriously. My speech act does not appear to change once I wonder whether I should change my belief in the face of disagreement.

14 One might raise here the possibility of treating it like neither of these options, but I do not know what the alternative would look like.

15 Jesper Kallestrup raised the alternative objection that Bar-On could treat the post-testimony case as being fully based on evidence, with the previous avowal being part of the evidential basis. This would still be incompatible with the authority that one has when one self-ascribes after considering the testimony.

4. CONCLUSION

I argued that mental state disagreements are cases in which the nature of self-knowledge affects my rational response to the disagreement. Given this feature, any account of self-knowledge has to be able to explain how self-knowledge influences what I ought to do. I proposed that accounts starting with self-knowledge as fully describable by linguistic practice cannot do that. Therefore setting up the problem according to the linguistic view seems to be misguided. Instead, we ought to explain self-knowledge as a phenomenon on the level of belief and belief formation. If this is correct, then the neo-expressivist project with a sole focus on our linguistic practice should be abandoned and rethought as something starting from a hybrid view, instead of the linguistic view.¹⁶

REFERENCES

- Bar-On, D. 2004. *Speaking My Mind*. Oxford: Oxford University Press.
- 2011. 'Neo-Expressivism: Avowals' Security and Privileged Self-Knowledge.' In A. Hatzimoyisis (ed.), *Self-Knowledge*, pp. 189–201. Oxford: Oxford University Press.
- Bar-On, D. and Sias, J. 2013. 'Varieties of Expressivism.' *Philosophy Compass*, 8(8): 699–713.
- Brueckner, A. 2011. 'Neo-Expressivism.' In A. Hatzimoyisis (ed.), *Self-Knowledge*, pp. 170–88. Oxford: Oxford University Press.
- Byrne, A. 2005. 'Introspection.' *Philosophical Topics*, 33(1): 79–104.
- Carruthers, P. 2011. *The Opacity of Mind: An Integrative Theory of Self-Knowledge*. Oxford: Oxford University Press.
- Cassam, Q. 2014. *Self-Knowledge for Humans*. Oxford: Oxford University Press.
- Finkelstein, D. 2003. *Expression and the Inner*. Cambridge, MA: Harvard University Press.
- Ryle, G. 1949 [1984]. *The Concept of Mind*. Chicago, IL: University of Chicago Press.
- Snowdon, P. 2012. 'How to Think about Phenomenal Self-Knowledge.' In A. Coliva (ed.), *The Self and Self-Knowledge*, pp. 243–62. Oxford: Oxford University Press.
- Williamson, T. 2000. *Knowledge and its Limits*. Oxford: Oxford University Press.
- Wright, C. 1989. 'Wittgenstein's Later Philosophy Of Mind: Sensation, Privacy, and Intention.' *Journal of Philosophy*, 86: 622–43.
- 1998. 'Self-Knowledge: The Wittgensteinian Legacy.' In C. Wright, B. Smith and C. Macdonald (eds), *Knowing Our Minds*, pp. 13–45. Oxford: Oxford University Press.
- 2001. *Rails to Infinity: Essays on Themes from Wittgenstein's Philosophical Investigations*. Cambridge, MA: Harvard University Press.
- 2015. Self-knowledge: The Reality of Privileged Access. In S. C. Goldberg (ed.), *Externalism, Self-Knowledge, and Skepticism*, pp. 49–74. Cambridge: Cambridge University Press.

LUKAS SCHWENGERER is a postdoctoral researcher at University Duisburg-Essen. He wrote his PhD on a transparency account of self-knowledge and works primarily in epistemology and philosophy of mind.

16 Special thanks to Michel Croce, Giada Fratantonio, Matt Jope, Jesper Kallestrup, Aidan McGlynn, and Kegan Shaw for feedback on earlier versions of this paper. Further thanks to Quassim Cassam who responded to a shorter version of the paper at MindGrad2017 at the University of Warwick and the reviewers for this journal for helpful comments. This article was written while I was funded by a UK Arts and Humanities Research Council doctoral studentship.