Synthese
https://doi.org/10.1007/s11229-022-03634-w

ORIGINAL RESEARCH

The transparency of expressivism

Wolfgang Freitag[1] · **Felix Bräuer**[2]

Received: 17 August 2021 / Accepted: 21 February 2022

Abstract
The paper argues that Gareth Evans' argument for transparent self-knowledge is based on a conflation of doxastic transparency with ascriptive transparency. *Doxastic* transparency means that belief about one's own doxastic state, e.g., the belief that one thinks that it will rain, can be warranted by ordinary empirical observation, e.g., of the weather. In contrast, *ascriptive* transparency says that self-*ascriptions* of belief, e.g., "I believe it will rain", can be warranted by such observation. We first show that the thesis of doxastic transparency is ill-motivated and then offer a non-epistemic interpretation of ascriptive transparency by reference to the theory of *explicit expressive acts*: "I think it will rain" requires attendance to the weather because the utterance expresses a belief about the weather, not about ourselves. This will allow us to avoid what is often called "the puzzle of transparent self-knowledge" while remaining faithful to Evans' linguistic observations.

Keywords Doxastic transparency · Ascriptive transparency · Expressivism · Gareth Evans · Wittgenstein

Gareth Evans (1982, 225) famously claimed that doxastic self-knowledge, knowledge of one's own beliefs, can be, and often is, grounded in the observation of external facts: whether we believe that it will rain can be found out by looking at the sky. And, to take Evans' own example, we are able to determine whether we believe that there is going to be a third world war by analyzing present politics. We can gain doxastic self-knowledge not, or not only, by introspection and other self-directed epistemic means, but (also) by looking at the world. This is often referred to as the "transparency" of self-knowledge.

✉ Felix Bräuer
felix.braeuer@phil.uni-mannheim.de

Wolfgang Freitag
wolfgang.freitag@phil.uni-mannheim.de

1 Lehrstuhl Philosophie I, Universität Mannheim, L 9, 5 – Raum 004, 68161 Mannheim, Germany

2 Lehrstuhl Philosophie I, Universität Mannheim, L 9, 5 – Raum 007, 68161 Mannheim, Germany

Published online: 11 April 2022

The transparency idea is very attractive, as it provides a countermodel to the introspection conception of self-knowledge and thus promises to clear up one of the great mysteries of the mind. Yet transparency gives rise to a serious puzzle of its own: how is it possible that answering a question about the external world (whether *p*) can yield knowledge about one's own beliefs? After all, *p* and *I believe that p* are logically independent propositions. In particular, the fact that it will rain, in and of itself, says nothing about whether we also *believe* this. Hence, it is unclear how belief in the former proposition should be able to serve as a basis for knowledge of the latter. A number of authors have attempted to meet the challenge, yet it is fair to say that, despite all the efforts, no fully convincing explanation has emerged so far.[1]

In light of this enduring problem, we will reevaluate the matter by having a closer look at the phenomenon of transparency itself. We will claim that Evans and the discussion following him have misidentified the phenomenon in need of explanation. The idea of transparent self-knowledge rests on a conflation of two different forms of transparency. *Doxastic transparency* claims that self-*belief* can be grounded in extrospection. Therefore, doxastic transparency gives rise to the transparency puzzle. Yet Evans' claim of doxastic transparency derives from a subtly different observation: *saying* "I believe that it will rain" can be warranted by looking at the weather alone. Call this latter version *ascriptive transparency*. It will be shown that, on a plausible reconstruction of Evans, doxastic transparency is inferred from ascriptive transparency by means of a highly contested and ill-motivated premise and is therefore itself unfounded. It will emerge that Evans is right in identifying ascriptive transparency as an interesting phenomenon, but wrong in deriving doxastic transparency from it. Thus, we can keep Evans' phenomenological insights, but avoid the puzzle of transparent self-knowledge.

The paper proceeds as follows. After having disentangled doxastic and ascriptive transparency in Sect. 1, we will provide an account of ascriptive transparency in Sect. 2. We do so by reference to a theory of expressive acts: an affirmative utterance of "I believe that it will rain" can be warranted by attending to the weather because, typically, this utterance does not express a self-directed belief at all, but rather a belief about the weather.[2] Thus, ascriptive transparency receives a *linguistic*, not an epistemological, explanation. We will then consider some objections in Sect. 3 and conclude with a brief summary in Sect. 4.

[1] According to inferentialism (cf., e.g., Byrne, 2011, 2018; Barz, 2019), the transition from *p* to *I believe that p* is accomplished by some kind of *inference*. And reflectivism (e.g., Boyle, 2009; 2011; 2019; see also Moran 2001, 2003) claims that the transition is accomplished by some kind of *reflection* on one's belief that *p*. Against inferentialism it has been argued that the second-order belief doesn't come out as first-person rational and hence isn't a genuine item of self-knowledge (cf. Boyle, 2011, 2019; see also Paul, 2014; Bar-On, 2015; Cassam, 2017). And reflectivism has been criticized for being a disguised introspectivist account of self-knowledge and thus failing to explain how transparent self-knowledge might be possible (cf., e.g., Barz, 2019; Paul, 2014).

[2] We say "typically" because we want to allow that affirmative utterances of the form "I believe that *p*" *can* also be used to report and express a self-directed belief. We discuss this ("descriptivist" and "non-explicit expressive") use in §§ 2 and 3 below.

1 Two kinds of transparency

Evans describes the transparency phenomenon in the following passage. (For the purpose of future reference, we label the two sentences.)

> [A] If someone asks me 'Do you think there is going to be a third world war?', I must attend, in answering him, to precisely the same phenomena as I would attend to if I were answering the question 'Will there be a third world war?' [B] I get myself in a position to answer the question whether I believe that *p* by putting into operation whatever procedure I have for answering the question whether *p*. (Evans, 1982, 225)

In [A], Evans compares two questions, "Do you think that *p*?" and "*p*?", with respect to the phenomena we must attend to in order to give epistemically warranted responses to those questions.[3] He claims that, in order to give such responses, we must merely focus on the state of current politics, if *p* concerns the third world war, and on the weather, respectively, if *p* states that it will rain. In short, Evans claims that we must consider the state of the world, not our state of mind. Assuming, with Evans, that a positive answer to "Do you think that *p*?" has the form "I think that *p*", [A] is concerned with the epistemic warrant of doxastic self-ascriptions, and hence with the epistemic rationality of *linguistic acts*: in order to be warranted in *affirmatively uttering* "I think there is going to be a third world war", one must attend to those facts whose consideration might warrant the affirmation of "There is going to be a third world war". Since [A] describes a special feature of doxastic self-*ascriptions*, it is concerned with what we call *ascriptive transparency*: warranted self-ascription of a belief that *p* requires attendance only to the *p*-justifying phenomena.

Let us have a closer look at the transition from [A] to [B]. We may ignore the unproblematic lexical switch from "think" to "believe": evidently, "think" is here used as a synonym for "believe", i.e., the holding true of a proposition.[4] Also, we don't want to quibble over the transition from the second person ("Do *you* think") to the first person ("whether *I* believe"): in the question "Do you think …?" the pronoun refers to the addressee of the question, who thereupon reflects upon it first person–style. The important step in Evans' transition from [A] to [B] consists in the change from a *direct* question ("Do you think that *p*?") to an *indirect* question ("whether I believe that *p*"), as this marks a switch in one's epistemic concern: it turns a question concerning the self-*ascription* of belief—the warrant for *saying* "I believe"—into a question about self-*belief*—the warrant for *believing that* I believe. In [B], Evans states that self-*belief*, the belief with the content *I believe that p*, can be based on the very same reasons the belief that *p* itself can be based on. Call this the thesis of *doxastic* transparency.

[3] Here and later, we assume that the sentence "*p*" describes a simple worldly proposition, such as "It will rain" or "There will be a third world war". In particular, we assume that it does not contain a first-person intentional phrase, such as "I think that …" or "I wish that …".

[4] To be clear, we don't assume that "think" and "believe" are interchangeable in all contexts (for discussion, see Roberts, 2021), yet we think it is fair to say that Evans uses them interchangeably in the context at hand. In what follows, we will mostly use the believe-locution.

[A] is concerned with the rationality of self-*ascriptions* and has—as such—nothing to do with self-belief or self-knowledge. It is only doxastic transparency which captures Evans' claim that self-knowledge is, at least sometimes, transparent. Thus [B] is required to state the idea of transparent self-knowledge. Yet although [A] and [B] represent logically independent theses, Evans moves from the one to the other without further argument. We must therefore ask: what licenses the transition from [A] to [B]?

Plausibly, Evans' (implicit) reasoning relies on the following auxiliary principles:

(AP1) An affirmative utterance of the form "I believe that p" is the assertion of, and hence the expression of a belief in, the proposition *that I believe that p*.
(AP2) Whatever epistemically licenses an affirmative utterance also licenses the belief expressed by that utterance.

Based on these auxiliary principles, Evans is then able to argue as follows: By (AP1), the speaker, with "I believe that p", asserts and expresses the belief that she believes that p. For example, with "I believe that it will rain", the speaker asserts and expresses the belief that she believes that it will rain. By (AP2), the epistemic warrant for the utterance is also the warrant for the (higher-order) belief expressed by means of the utterance, viz., the speaker's belief that she believes it will rain. Hence, if *saying* "I believe that it will rain" is warranted by extrospection, then so is the speaker's *belief* that she believes that it will rain. Thus, given (AP1), (AP2), and ascriptive transparency, doxastic transparency follows. In this way, the two auxiliary principles permit the move from the thesis that the self-*ascription* of belief can be based on extrospection (in [A]) to the thesis that self-*belief* can be based on extrospection (in [B]).

We will assume that (AP2) is correct, at least for the cases at hand, but argue that (AP1) is highly problematic and actually untenable in the present context. (AP1) is an instance of the general *affirmation–belief principle* stating that the affirmative utterance of "φ" is the assertion and hence expression of the belief that φ. With "It will rain", the speaker asserts and expresses the belief that it will rain, and with "There will be a third world war", she asserts and expresses the belief that international politics will take a disastrous turn. Thus, if the affirmation–belief principle holds unrestrictedly, with "I believe that it will rain", the speaker expresses the belief that she believes it will rain, and with "I believe that there will be a third world war", the belief that she believes in a catastrophic development of politics. Thus, the affirmation–belief principle might explain why (AP1) appears initially plausible and why Evans blithely moves from [A] to [B].

Yet even though the affirmation–belief principle is plausible for many cases, it is arguably incorrect precisely for those cases we are interested in here: self-ascriptions of mental states in general, and self-ascriptions of belief in particular. Amongst many others, Wittgenstein, in a famous passage from the *Philosophical Investigations*, states that doxastic self-ascriptions are not self-reports, as their surface grammar would suggest, but hedged assertions of the embedded proposition itself. He thus contrasts statements of the form "I believe that p" with "*He* believes that p" and "I believ*ed* that p", which are reports about *another* person's present doxastic state and the speaker's own *past* one, respectively. Wittgenstein's claim is captured by

> [W1] "*I* believe that *p*" is *not* an assertion about one's belief that *p* at all, but a "hesitant assertion" of *p* itself.[5]

Yet if [W1] is correct, (AP1) arguably isn't: if "I believe that *p*" is used to (hesitantly) assert *p*, it expresses the belief in the embedded proposition *p*, not the (higher-order) belief in the proposition that one *believes* that *p*: "I believe that it will rain" and "I believe that there will be a third world war" would not be used to express higher-order beliefs about the speaker's beliefs. Rather, the speaker would use them to (hesitantly) assert and express her *first-order* beliefs concerning the future of both the weather and the world.

[W1] blocks the transition from [A] to [B]: if "I believe that *p*" is used by the speaker to (hesitantly) express a first-order belief about *p*, then it tells us nothing about which higher-order beliefs the speaker might have. Indeed, [W1] seems to directly contradict [B]. After all, [W1] strongly suggests that "I believe that *p*" does not express a higher-order belief in the first place; hence the mental state expressed wouldn't even be a candidate for *self*-knowledge. It would be a candidate for knowledge concerning the state of the world, not concerning one's state of mind.

In defence of Evans' transition from [A] to [B], one might try to undermine Wittgenstein's view on doxastic self-ascriptions and argue in favour of (AP1). Yet, this move isn't open to Evans, as he even derives his transparency claims from Wittgenstein's views. Evans begins his famous discussion of transparency with the following quote from Wittgenstein:

> [W2] If a man says to me, *looking at the sky*, 'I think it will rain, therefore I exist,' I do not understand him. (originally in Coope et al., 1970, 21; quoted from Evans, 1982, 225; Evans' italics)

Evans comments as follows: "I think Wittgenstein was trying to undermine the temptation to adopt a Cartesian position, by forcing us to look more closely at the nature of our knowledge of our own mental properties, and, in particular, by forcing us to abandon the idea that it always involves an *inward* glance at the states and doings of something to which only the person himself has access" (Evans, 1982, 225). Evans thus takes Wittgenstein to propose an anti-Cartesian epistemology of the mind and to identify the phenomenon of doxastic transparency as described in [B]. So, Evans does not defend his claim of doxastic transparency *against* Wittgenstein, but ironically takes it *from* Wittgenstein.

But couldn't Evans simply use his interpretation of [W2] for his purposes and ignore [W1]? We think that this isn't a live option. For one thing, interpreting [W2] in terms of [B] is clearly at odds with a plausible reconstruction of the remark. In [W2], Wittgenstein is apparently concerned with the validity of the inference of "I exist" from "I think it will rain" if the latter is uttered in the context of extrospection ("looking

[5] Wittgenstein (1953), part II x, 190–192 (cf. Wittgenstein, 1980a, §§ 471–478; 1980b, §§ 279–283). Similar views have been expressed by a number of other thinkers, e.g., by Urmson (1952), 483–484; Benveniste (1974, 228); Giorgi & Pianesi 2005, 112; Krifka 2014, 81–82; Kemmerling 2017; Henning 2018. For the purpose of this paper, we use "hedged assertion" and "hesitant assertion" interchangeably; they both refer to forms of weakened assertoric force.

at the sky").[6] Yet questions of validity concern the logical relation between premises and conclusion alone and are hence independent of the justification (and even the truth) of the premises. In other words: if Evans' epistemic interpretation in terms of [B] were correct and [W2] were meant to show that the premise expresses a self-belief warranted by extrospection, then not the inference from "I think" to "I exist" but rather Wittgenstein's critique thereof would be incomprehensible. Furthermore, Evans' interpretation would put [W2] in straightforward conflict with Wittgenstein's denial of (AP1) in [W1]. Now, [W2] is not only consistent with [W1], it can even be explained in terms of [W1]. If "I think it will rain" is but the hesitant assertion of the embedded proposition *that it will rain*, then "I exist", as a statement about the speaker's own existence, cannot be inferred from said assertion—just as "I exist" cannot be inferred from "It will rain".

The thesis of doxastic transparency [B] is in conflict with both [W1] and [W2] and thus in conflict with Wittgenstein's ideas. Since Evans' reasons for accepting doxastic transparency are actually based on an interpretation of Wittgenstein's remark [W2], and since he does not provide any further, independent support, he gives us no reason to believe in *doxastic* transparency. We thus take the thesis of doxastic transparency to be unfounded.

What about ascriptive transparency, described in [A]? Evans makes a second comment about [W2]: "The crucial point is the one I have italicized: in making a *self-ascription of belief*, one's eyes are, so to speak, or occasionally literally, directed outward—upon the world" (Evans, 1982, 225; our italics). In this second comment, Evans does not refer to self-knowledge at all, but to self-*ascription* of belief. Thus, we take it, Evans identifies the "crucial point" of [W2] to be that of *ascriptive* transparency (which is then explicitly stated in the sentence that immediately follows, [A]): in order to self-ascribe a doxastic state, the agent does not have to look inward, at her state of mind; she simply needs to look at the state of the world.

Note that [A] is fully in tune with both passages from Wittgenstein, even more so as [W1] does not only explain [W2] but also the phenomenon of *ascriptive* transparency. If "I think it will rain" is the hesitant assertion of the proposition *that it will rain*, its warrant is naturally the same as that of the assertion "It will rain"—e.g., the perception of dark clouds on the horizon.

It appears therefore that Evans' claim of doxastic transparency is the result of, first, an important insight and, second, a subtle confusion. The insight is that self-ascriptions of belief can be grounded in the observation of external facts. This is the phenomenon of ascriptive transparency. Yet, perhaps due to the acceptance of an unrestricted affirmation–belief principle, he mistakes *ascriptive* transparency for *doxastic* transparency, thus creating the impression that transparency also applies to *self-knowledge*. Once we have disentangled the two kinds of transparency at issue, it emerges that Evans does not provide any good reason for assuming that there is such a thing as doxastic transparency. And since doxastic transparency leaves us with the

[6] One might be tempted by another interpretation, according to which [W2] draws attention to the fact that the inference to "I exist" has no use in ordinary language. Yet that temptation must be resisted, because it would render the reference to the speaker's extrospection futile, while Wittgenstein evidently thinks it to be crucial for the case. Observe that Evans clearly does not yield to this temptation.

aforementioned transparency puzzle, there are actually good reasons to be sceptical of it.

It emerges that the transparency debate has focused on the wrong explanandum. It should be concerned with the phenomenon for which Evans actually makes a convincing case: *ascriptive* transparency. To this we will now turn.

2 Ascriptive transparency and explicit expressives

The real interest in Evans' passage resides in the claim that there is ascriptive transparency, and hence in [A]: what would account for the fact that *self-ascriptions* of belief (e.g., "I believe that it will rain") can be justified by reference to external facts (e.g., the colour of the sky)? We aim at answering this question by providing an explanation in terms of explicitly expressive acts.[7] To introduce the idea, let us start from scratch.

Linguistic acts typically have both an illocutionary and an expressive dimension.[8] An utterance of "Come to my party!" is a request used by the speaker to express a desire.[9] An utterance of "I'll come to your party" constitutes a promise, and it expresses the speaker's intention to show up at the party. And the affirmation of "It will rain", finally, is both an assertion about the weather and the expression of the corresponding belief. We will here focus on the *expressive* dimension of speech and largely ignore its illocutionary dimension.[10]

Expressive acts have expressive contents. The expressive content of an utterance of "Come to my party!" is the speaker's corresponding *desire*. And the expressive content of the answer "I'll come to your party" is the corresponding *intention*. The expressive content of "It will rain", lastly, is the speaker's corresponding *belief*. That a speaker S expresses content C can be represented in a function notation of the form $\mathrm{Exp}_S(C)$. (We

[7] Recently, also an alternative, evidentialist analysis of first-person belief ascriptions has been suggested. (Thanks to an anonymous referee for pointing that out.) According to, for example, Koev (2021), a speaker who affirmatively utters "I believe that p" can do so to make a primary assertion to the effect that p and a secondary assertion to the effect that she *believes* that p, which is used to provide evidence for the truth of the primary assertion (cf. Koev, 2021, 115–116). While it might be tempting to explore the extent to which ascriptive transparency could also be understood as derivable from an evidential use of "I believe that p", doing so would be well beyond the scope of the present paper.

[8] We won't discuss in this paper the possible relation between the notions of 'expression' and 'expressing a mental state' to the notion of 'commitment' (as recently discussed, e.g., by Shapiro, 2020). Our argument turns less on the notion of expression itself than on the content expressed (or committed to). Also, note that a given utterance *need not* possess both an illocutionary and an expressive dimension (although this may typically be the case). A person might, for example, express a desire without committing a directive speech act (as in "I would like you to study law, but I also want you to decide autonomously").

[9] We distinguish between a *speaker's* expressing a *mental state* and a *sentence's* expressing a *proposition*. Compare Bar-On (2004), who, following Sellars (1969), distinguishes "s-expression" (on the sentence level) and "a-expression" (on the utterance or act level).

[10] We conceive of an *expressive* act as a linguistic act—an utterance by which the speaker expresses that she is currently in a certain mental state (cf., e.g., Bach & Harnish, 1979; Bar-On, 2004; Finkelstein, 2003; Green, 2007). We don't conceive of an expressive act as an *illocutionary* act. In particular, an expressive act, in our sense, cannot be subsumed under the category of what John Searle (1976) calls "expressives". We cannot here discuss possible relations to the expressivist theories developed by, for example, Potts (2007) and Gutzmann (2015).

will usually omit indexing to the subject S.) Since C is the subject's mental state, and since a mental state—a wish, an intention, a belief—may itself have a representational content, we will distinguish within the content C between the (subjective) *mental force* and the subordinate *mental state content*, where subjective mental force is determined by the psychological state expressed. The mental state content of "Come to my party!" is *that you come to my party*, and that of "It will rain" is *that it will rain*. Using ψ as a schematic letter indicating mental force in general, "Exp($\psi(p)$)" represents S's expressing a mental state ψ with the (propositional) content p. For example, if δ stands for the specific attitude of desire, the expressive act performed with an utterance of "Come to my party!" is represented by Exp(δ(You will come to my party)), which says that the speaker expresses a desire that the addressee come to her party. And if β stands for belief, the expressive act performed with an utterance of "It will rain" is represented by Exp(β(It will rain)).[11]

Up to this point our claims are rather straightforward and, apart perhaps from the terminology, fairly familiar. Now to the interesting bit. Utterances of "I'll come to your party", of "Come to my party!", and of "It will rain" are *non-explicit* expressive acts in that the speaker expresses an intention, a wish, and a belief, respectively, *without mentioning* the mental state expressed and the subject possessing it. Our central claim is that there are also *explicit* expressives, in which mental state and subject *are* mentioned. What "Come to my party!" *displays* (by syntactical form), namely that the speaker expresses a wish concerning the addressee's behaviour, is *verbally articulated* in an utterance of "I wish that you come to my party".[12] Similarly, we claim, with "I intend to come to your party", the speaker expresses the corresponding intention, and with "I think that p" the belief that p. Explicit expressives are typically of the form

$$(\mathrm{I} - \psi)\ \mathrm{I}\psi \text{ that } p,$$

where the dominant verb ψ refers to a psychological state and is in the first-person (non-progressive) present-tense indicative active.[13]

[11] Although expressive acts are neither true nor false, they can be correct or incorrect in an important sense. The expressive act performed by uttering "It will rain" is *expressively correct* iff the speaker indeed believes that it will rain. And the expressive act performed by uttering "I am sorry for being late" is *expressively correct* iff the speaker is indeed sorry for being late (cf. Kaplan, 1999, 9). More generally, S's expressive act with the expressive form Exp($\psi(p)$) is expressively correct iff S indeed ψ's that p. We hold that an expressive act *need* not be expressively correct (e.g., if the speaker is insincere or deluded). (For other non-factive notions of expression, see also Austin, 1962, lect. IV; Searle, 1969, 65; Harnish ,1976; Kemmerling, 2002; Davis, 2003; Eriksson, 2010; Marsili, 2016. But see Rosenthal, 1986, 1993; Malcolm, 1991; Green, 2007; 2008; and Williams, 2013, for factive theories of (self-)expression).

[12] Recent expressivists, such as Finkelstein (2003; 2012) and Bar-on (2004), also claim that, with the self-ascription of a mental state, the speaker expresses that very mental state. In contrast to us, however, they hold that the speaker additionally expresses *belief* in that very mental state. They use this dual expression thesis to solve the puzzle of doxastic transparency where we deny that there is a puzzle to be solved in the first place.

[13] Note that mental force might also be encoded in plural versions of (I-ψ) and sentence adverbs such as "hopefully" and "unfortunately". For a discussion of sentence adverbs in relation to the theory of explicit expressives and to Frege's notion of colouring, see Freitag (2014).

The traditional descriptivist analysis would hold that with an affirmative utterance of this form the speaker *describes* her mental state: she *asserts* that she ψ's that p and, since assertion is at the same time an expression of belief, *expresses a belief* in that very proposition. According to the descriptivist analysis, the affirmation of "I think that p" is the expression of the second-order belief *that I think that p* and thus has the expressive form Exp($\beta(\beta(p))$). And the content of that second-order belief, its mental state content, is the proposition that she, the speaker, is in the first-order mental state of believing *that p*. Premise (AP1), and thus Evans' claim of doxastic transparency, rests on the descriptivist analysis of self-ascriptions of mental states.[14]

The theory of explicit expressives rejects this analysis. It maintains that, with an utterance of the form (I-ψ), the speaker typically does not express the self-directed belief that she ψ's that p, but rather the first-order mental state of ψ'ing that p itself. The first-person prefix "I ψ that" does not contribute to the content of the mental state expressed by the speaker: the mental state predicate "ψ" indicates mental force, the expressed mental state alone. And the first-person pronoun is not something the speaker has a belief about, but rather refers to the subject of that mental state. The content of the mental state expressed, the mental state content, is therefore p alone. The theory of explicit expressives claims, in other words, that the expressive form of utterances of the form (I-ψ) is Exp($\psi(p)$). The expressive form of "I think that p" is therefore Exp($\beta(p)$), not Exp($\beta(\beta(p))$).

We are now in a position to describe the relation between the semantic content of a sentence (the proposition expressed by the sentence) and the representational content of the mental state expressed. In *non-explicit* expressives both contents coincide: the semantic content is identical to the content of the mental state expressed. In an utterance of "It will rain", for example, the semantic content of the sentence, *that it will rain*, is identical to the content of the mental state expressed, i.e., the belief *that it will rain*. In *explicit* expressives, however, the semantic content and the mental state content come apart: the *semantic content* of "I think that it will rain" is *that I think that it will rain*; the *mental state content*, on the other hand, is the proposition embedded in the semantic content, i.e., in the present case, *that it will rain*. The expressive content of "I think that it will rain" is identical to the expressive content of "It will rain": Exp(β(It will rain)). The very same expressive act can hence assume two different linguistic forms—that of a non-explicit and of an explicit expressive.

The difference between the non-explicit and the explicit can also be made with respect to questions. The question "Will it rain?" concerns the weather. And in communicative contexts, we pose this question by asking the addressee A for her opinion. In answering "It will rain" or "I think it will rain", A expresses a belief about the weather. Assuming A's sincerity, we can infer what A believes. In some cases, e.g., in exams, that's all we want with such a question. More often, however, we pose the question to find out about the weather. And to the degree that we assume A to be competent, the fact that A believes that it will rain will also influence our own beliefs about the weather. Thus, questions are, or can be, Janus-faced, much like other speech

[14] Amongst others, Fleming (1955), Austin (1962), Tomberlin (1968), Green (1970), Searle (1979), Rosenthal (1986, 1993, 1995a, 1995b, 2010), Kaplan (1999), and Green (2009) propose descriptivist analyses of serious and competent utterances of the form (I-ψ).

acts are. Posed to A, the question "Will it rain?" is directed at some subject matter (the weather) and in addition concerned with the addressee's mental states.

The question "Will it rain?" is *non-explicit* with respect to what psychological state of A is to be expressed. Yet, we hold, there are also questions which are explicit in this respect. With "Do you think it will rain?", we pose the same question about the weather, only that we now address A and what we demand of her in a verbally explicit manner. The doxastic prefix "Do you think that" refers in a verbally explicit way to the addressee and the mental state whose expression is demanded. Again, the analysis can be generalized. A question of the form

$$(\text{Do} - \text{you} - \psi?)\text{Do you } \psi \text{ that } p?,$$

with ψ referring to some psychological state, is an explicit question in this sense. With "Do you wish that p?", the speaker wants the addressee to express her wishes regarding p, and with "Do you intend to come?" she demands the addressee to voice her intentions. With "Do you think that it will rain?", finally, S asks A for her opinion about the weather, not for her opinion about herself. That is, the speaker does not inquire *about* the addressee's desires, intentions, and beliefs but she demands *of* the addressee to *express* her wishes, intentions, and beliefs. In each case, the addressee is expected to *express* a first-order mental state rather than to report or describe it. The addressee's proper response is to express her attitude towards p by either an explicit or a non-explicit expressive. To the question "Do you think it will rain?," A might react either with "It will rain" or with "I think that it will rain", and give essentially the same answer in both cases.

The theory of explicit expressive questions explains ascriptive transparency. The questions "p?" and "Do you think that p?" ask (for) the same thing in different linguistic forms: they both concern p and the addressee's doxastic state with respect to p. This fully accounts for Evans' observation in [A] that "if someone asks me 'Do you think there is going to be a third world war?', I must attend, in answering him, to precisely the same phenomena as I would attend to if I were answering the question 'Will there be a third world war?'". In answering either question, one must attend to the same phenomena because both questions ask the same: whether a third world war is upon us and whether the addressee believes so.

Alternatively, we may describe our solution with respect to the congruence, not of the questions, but of the answers. Given that the *expression* of a belief is epistemically warranted if the belief itself is epistemically warranted, i.e., given (AP2), and given that the belief that p is warranted by considerations about p alone, A's response "I think that there will be a third world war" requires, just like the response "There will be a third world war", only the consideration of present politics. Ascriptive transparency, as described in [A], is explained by the linguistic theory of explicit expressives. No reference to the epistemic particularities of self-beliefs is required and no puzzle concerning self-knowledge needs to be solved.[15]

[15] We may, of course, generalize to arbitrary self-ascriptions of propositional attitudes: the theory of explicit expressives explains also why, say, the question "Do you want red or white wine?" requires exploring the menu, not the mind.

3 Objections and replies

Let us address three possible objections. *First*, one might object that our analysis was in part motivated by a critique of (AP1), as found in Wittgenstein and others, but that this critique is based on an analysis of self-ascriptions of doxastic states which is itself at odds with ours. As described above, [W1] claims that self-ascriptions of beliefs are (hesitant) assertions of the embedded proposition. There is no mention of the idea that a doxastic self-ascription is the expression of a first-order belief.

We hold that the conflict is apparent only: the theory of explicit expressives can and should be regarded as a natural generalization of Wittgenstein's own views, once they are properly understood. Wittgenstein himself proposes that affirmative utterances of the form (I-ψ) typically are explicit expressives (without of course using this term). He would construe "I intend to come to your party" as the expression of the speaker's intention, and "I want red wine" as the expression of a desire.[16] And although he does not say so outright, we suggest that he would also accept that "I think that p" is the explicit expression of a first-order belief, because a good case can be made for the claim that, typically, to explicitly express a belief is to implicate the affirmation of the embedded content (see Freitag, 2018 for detailed discussion): an affirmative utterance of "I believe that it will rain" is usually best explained by the fact that the speaker intends to comment on the weather, without undertaking the communicative obligations that come with the outright assertion of "It will rain". If this is correct, Wittgenstein's view that "I think it will rain" is a hesitant assertion about the weather—the claim expressed in [W1]—is not in conflict with, but actually follows from, the theory of explicit expressives. We therefore think that Wittgenstein is best construed as embracing the theory of explicit expressives also for doxastic self-ascriptions.

Naturally, this construal would also explain Wittgenstein's complaint in [W2]: from "I think it will rain", used as an explicit expression of the belief that it will rain, the speaker can no more infer that she exists than from the non-explicit expressive "It will rain". The theory of explicit expressives therefore not only provides an explanation of ascriptive transparency, but also of Wittgenstein's remark from which it is derived.

Second, we have discussed the theory of explicit expressive acts only for *positive* self-ascriptions of mental states. But, it may be observed, a question such as "Do you think it will rain?" may require a negative answer. Looking at the sky and not coming to any conclusion, the addressee might respond with "I don't think it will rain", thereby leaving it open whether she believes that it won't rain or suspends judgement on the matter. Yet if the question is to be construed as an explicit question in our sense, then "I don't think it will rain" must also be understood as something like an explicit expressive, despite the fact that it does not have the form (I-ψ); it is, after all, a *negated* self-ascription of a mental state.

In response to this challenge, we suggest extending the theory of explicit expressives to negated self-ascriptions of mental states. We claim that there is *expressive denegation*: when we say "I don't ψ that p", we don't express the *belief* that we

[16] According to Wittgenstein, "I hope he'll come" is the expression of the speaker's hope (Wittgenstein, 1953, § 585; see also 1967, § 78), "I am in pain" is the expression of pain (Wittgenstein, 1953, §§ 244, 404; also 1958, 68), and "I expect a bang" is the expression of an expectation (Wittgenstein, 1967, § 53). Compare also Wittgenstein 1980a, §§ 469, 472, and 477.

don't ψ that p, rather we express that we are *not* in the state of ψ'ing p. For example, with "I don't want red wine", the speaker expresses indifference, and with "I don't intend to come to your party", indecision.[17] And if, in response to the question "Do you think it will rain?", *A* says, "I don't think it will rain", she does not make an assertion about her psychological condition. She does not express the *belief* that a belief that it will rain is absent, rather she expresses that very absence. She expresses the absence of belief in response to the fact that the evidence does not warrant belief concerning the development of the weather.[18]

Third, and finally, it might be objected that the present take on transparency surely cannot be the whole story. After all, statements of the form "I think that p" *need not* be used as explicit expressives; they can also be used as *non*-explicit expressives, as actual descriptions of one's mental states. More generally, utterances of the form "I ψ that p" *may* be used to report one's own mental state of ψ-ing and hence indeed to express the *belief* that one is ψ-ing that p. To give an example: Linda, who learns through a test that she harbors implicit biases, might say something like "I believe that women make worse political leaders than men" to report her newly-discovered attitudes, and thus to express a belief about her own attitudes, rather than to explicitly express the first-order attitude. And likewise, when someone asks Linda "Do you (really) think that p?" the questioner might not inquire about p but rather *about* Linda's beliefs. The inquirer then demands of Linda the expression, not of a belief about p, but of a belief about her own mental state.

We grant both the explicit and the non-explicit expressive use of utterances of the form (I-ψ), and of the corresponding questions (Do-you-ψ?). We think, however, that in these uses of "I believe that p" there is no ascriptive transparency. If the speaker is in one of these rare moods or situations in which she actually *reports* her first-order beliefs concerning p and hence expresses a *second*-order belief, her reasons cannot be p-related. Linda's reasons for the belief that she is prejudiced about women do not reside in her observations of female success in politics, but in some evidence concerning her own beliefs: she looks at her test results and thereby discovers her sexist attitudes. More generally, it seems that the non-explicit expressive utterance of "I ψ that p" cannot be warranted by reference to p alone, but needs evidence concerning the subject's mental state of ψ-ing that p. We maintain, therefore, that ascriptive transparency is bound to the explicit expressive use of self-ascriptions of mental states. So, there is no obstacle to explaining ascriptive transparency by reference to the theory of explicit expressives.

[17] We here ignore the phenomenon of neg-raising, according to which "I don't ψ that p" can be used and understood in the way of "I ψ that not-p". In our view, neg-raising is a frequent, but *pragmatic*, phenomenon. (For this assessment see also Bartsch, 1973; Horn, 2001; Gajewski, 2007; and Zeijlstra, 2018). And if neg-raising were a semantic phenomenon, negated self-ascriptions of mental states would not require special treatment.

[18] The notion of expressive denegation gives rise to a further question. While it seems fairly plausible that one can express a mental state, it is less clear what it means for a speaker to express its absence: the intentional object of expression is then not a positive mental state but the *absence* of such a positive mental state. There are several possible reactions, some of which are explored in Freitag & Yolcu (2021), where the idea of expressive denegation is discussed in more detail.

4 Conclusion

The aim of this paper was to dispel the mist surrounding Evans' transparency and thus to critically re-evaluate the basis of the contemporary debate on transparent self-knowledge. We have first shown that Evans refers to two different notions of transparency, which must be carefully set apart: ascriptive and doxastic transparency. We have further suggested that Evans mistakes the former for the latter and thus wrongly takes the arguments for ascriptive transparency to be arguments for doxastic transparency. Given this result, we have rejected Evans' claim of *doxastic* transparency and hence also the notion of transparent self-*knowledge*. This has allowed us to avoid the puzzle that notoriously plagues the debate on transparent self-knowledge. We have then set out to show that *ascriptive* transparency is due to the linguistic peculiarity of self-ascriptions, and not to an epistemic anomaly of self-belief. By developing the theory of explicit expressives, our aim was, to quote Wittgenstein once more, to "condense a cloud of philosophy into a drop of grammar" (1953, § 222). Looking at the results, we don't think that there is going to be a third world war, but we think it will rain.

Acknowledgements We thank the members of the research colloquium of the Chair for Theoretical Philosophy at the University of Mannheim (especially Nadja-Mira Yolcu) and two anonymous referees of this journal for helpful comments on previous drafts. Christopher von Bülow did the proofreading and editing.

Funding Open Access funding enabled and organized by Projekt DEAL. The research was generously supported by the Fritz Thyssen Stiftung.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

References

Austin, J. L. (1962). *How to do things with words*, 2nd (Edn). In J. O. Urmson & M. Sbisà (Eds.). Harvard University Press.

Bach, K., & Harnish, R. (1979). *Linguistic communication and speech acts*. MIT Press.

Bar-On, D. (2004). *Speaking my mind: Expression and self-knowledge*. Clarendon Press.

Bar-On, D. (2015). Transparency, expression, and self-knowledge. *Philosophical Explorations, 18*(2), 134–152.

Bartsch, R. (1973). Negative transportation' gibt es nicht. *Linguistische Berichte, 27*, 1–7.

Barz, W. (2019). The puzzle of transparency and how to solve it. *Canadian Journal of Philosophy, 49*(7), 916–935.
Benveniste, É. (1974). *Probleme der allgemeinen Sprachwissenschaft*. List-Verlag.
Boyle, M. (2009). Two kinds of self-knowledge. *Philosophy and Phenomenological Research, 78*(1), 133–164.
Boyle, M. (2011). Transparent self-knowledge. *Proceedings of the Aristotelian Society, 85*(1), 223–241.
Boyle, M. (2019). Transparency and reflection. *Canadian Journal of Philosophy, 49*(7), 1012–1039.
Byrne, A. (2011). Transparency, belief, intention. *Proceedings of the Aristotelian Society, 85*(1), 201–221.
Byrne, A. (2018). *Transparency and self-knowledge*. Oxford University Press.
Cassam, Q. (2017). What asymmetry? Knowledge of self, knowledge of others, and the inferentialist challenge. *Synthese, 194*(3), 723–741.
Coope, C., et al. (1970). *A Wittgenstein workbook*. Basil Blackwell.
Davis, W. (2003). *Meaning, expression, and thought*. Cambridge University Press.
Eriksson, J. (2010). Self-expression, expressiveness, and sincerity. *Acta Analytica, 25*, 71–79.
Evans, G. (1982). *The varieties of reference*. In J. McDowell (Eds.). Clarendon Press.
Finkelstein, D. H. (2012). "From transparency to expressivism". In G. Abel & J. Conant (Eds.), *Rethinking epistemology*, vol. 2 (Berlin Studies in Knowledge Research), (pp. 101–118). De Gruyter.
Finkelstein, D. H. (2003). *Expression and the inner*. Harvard University Press.
Fleming, B. N. (1955). On avowals. *The Philosophical Review, 64*(4), 614–625.
Freitag, W., & N.-M. Yolcu (2021). An expressivist solution of moorean paradoxes. *Synthese* (Online First), 1–24.
Freitag, W. (2014). Frege über, leider' und, gottlob'. In B. Reichardt & A. Samans (Eds.), *Freges Philosophie nach Frege* (pp. 161–174). Brill Mentis.
Freitag, W. (2018). Wittgenstein on 'I believe.' *Grazer Philosophische Studien, 95*(1), 54–69.
Gajewski, J. R. (2007). Neg-raising and polarity. *Linguistics and Philosophy, 30*(3), 289–328.
Giorgi, A., & F. Pianesi (2005). Credo (I Believe): Epistemicity and the syntactic representation of the speaker. *University of Venice Working Papers in Linguistics*, *15*, 105–152.
Green, M. S. (2007). *Self-expression*. Oxford University Press.
Green, M. S. (2008). Expression, indication and showing what's within. *Philosophical Studies, 137*(3), 389–398.
Green, M. S. (2009). Speech acts, the handicap principle and the expression of psychological states. *Mind & Language, 24*(2), 139–163.
Green, O. H. (1970). The expression of emotion. *Mind, 79*(316), 551–568.
Gutzmann, D. (2015). *Use-conditional meaning: Studies in multidimensional semantics*. Oxford University Press.
Harnish, R. (1976). Logical Form and implicature. In T. G. Bever, J. J. Katz, & D. T. Langendoen (Eds.), *An integrated theory of linguistic abilities* (pp. 313–392). The Harvester Press.
Henning, T. (2018). *From a rational point of view: How we represent subjective perspectives in practical discourse*. Oxford University Press.
Horn, L. R. (2001). *A natural history of negation (The David Hume Series)*. CSLI Publications.
Kaplan, D. (1999). "The meaning of 'Ouch' and 'Oops': Explorations in the theory of *meaning as use*", 2004 version. University of California, Los Angeles (unpublished manuscript).
Kemmerling, A. (2002). Expressing an intentional state. In G. Grewendorf & G. Meggle (Eds.), *Speech acts, mind and social reality: Discussions with John R. Searle* (pp. 83–91). Kluwer Academic Publishers.
Kemmerling, A. (2017). *Glauben Essay über einen Begriff*. Klostermann.
Koev, T. (2021). Parentheticality, assertion strength, and polarity. *Linguistics and Philosophy, 44*(1), 113–140.
Krifka, M. (2014). Embedding illocutionary acts. In T. Roeper & M. Speas (Eds.), *Recursion: Complexity in cognition* (pp. 59–87). Springer.
Malcolm, N. (1991). I believe that *p*. In E. Lepore & R. van Gulick (Eds.), *John Searle and his critics* (pp. 159–168). Basil Blackwell.
Marsili, N. (2016). Lying by promising: A study on insincere illocutionary acts. *International Review of Pragmatics, 8*, 271–313.
Paul, S. (2014). The transparency of mind. *Philosophy Compass, 9*(5), 295–303.
Potts, C. (2007). The expressive dimension. *Theoretical Linguistics, 33*(2), 165–198.
Roberts, T. d. H. (2021). *How to make believe: Inquisitivity, veridicality, and evidentiality in belief reports*. Doctoral dissertation. University of California at Santa Cruz.

Rosenthal, D. M. (1986). "Intentionality". *Midwest studies in philosophy*, vol. 10 (pp. 151–184). Reprinted in D. M. Rosenthal, *Consciousness and mind* (pp. 71–99). Oxford University Press, 2005.
Rosenthal, D. M. (1993). "Thinking that one thinks". In M. Davies & G. W. Humphreys (Eds.), *Consciousness: Psychological and philosophical essays* (pp. 197–223). Basil Blackwell. Reprinted in D. M. Rosenthal, *Consciousness and mind* (pp. 46–70), Oxford University Press, 2005.
Rosenthal, D. M. (1995a). Moore's paradox and consciousness. *Philosophical Perspectives, 9*, 313–333.
Rosenthal, D. M. (1995b). Self-knowledge and Moore's paradox. *Philosophical Studies, 77*(2–3), 195–209.
Rosenthal, D. M. (2010). Expressing one's mind. *Acta Analytica, 25*, 21–34.
Searle, J. R. (1969). *Speech acts: An essay in the philosophy of language*. Cambridge University Press.
Searle, J. R. (1976). A classification of illocutionary acts. *Language in Society, 5*, 1–23.
Searle, J. R. (1979). *Expression and meaning*. Cambridge University Press.
Sellars, W. (1969). Language as thought and as communication. *Philosophy and Phenomenological Research, 29*(4), 506–527.
Shapiro, L. (2020). Commitment accounts of assertion. In S. Goldberg (Ed.), *The Oxford handbook of assertion* (pp. 73–97). Oxford University Press.
Tomberlin, J. E. (1968). The expression theory of avowals. *Philosophy and Phenomenological Research, 29*(1), 91–96.
Urmson, J. O. (1952). Parenthetical verbs. *Mind, 61*, 480–496.
Williams, J. N. (2013). Moore's paradox and the priority of belief thesis. *Philosophical Studies, 165*, 1117–1138.
Wittgenstein, L. (1967). *Zettel* 2nd (Edn). In G. E. M. Anscombe & G. H. von Wright (Eds.), Translated by G. E. M. Anscombe. Blackwell.
Wittgenstein, L. (1980b). *Remarks on the Philosophy of Psychology*, vol. II. In: Edited by G. H. von Wright & H. Nyman. Chicago, Ill.: University of Chicago Press.
Wittgenstein, L. (1980a). *Remarks on the Philosophy of Psychology*. In G. E. M. Anscombe & G. H. von Wright (Eds.), vol. I. Blackwell.
Wittgenstein, L. (1953). *Philosophical investigations* (2nd ed.). Blackwell.
Wittgenstein, L. (1958). *The blue and brown books*. Harper & Row.
Zeijlstra, H. (2018). Does neg-raising involve neg-raising? *Topoi, 37*, 417–433.