International Philosophical Quarterly
doi:
Online First:

Transparent Self-Knowledge of Attitudes and Emotions: A Davidsonian Attempt

Ning Fan

ABSTRACT: In *Authority and Estrangement*, Richard Moran provides a fascinating account of how we know what we believe that he calls the "transparency account." This account relies on the transparency relation between the question of whether we believe that *p* and the question of whether *p* is true. That is, we can consider the former by considering the grounds for the latter. But Moran's account has been criticized by David Finkelstein, who argues that it fails to explain how we know our attitudes and emotions more generally. The aim of this paper is to show how Moran's transparency account can be extended to meet this criticism by modifying it, using insights from Davidson's view on attitudes and emotions.

1. INTRODUCTION

IN DETERMINING WHETHER THERE IS AN APPLE TREE in my backyard, I can open the window and look at my backyard to see whether there is an apple tree. What I do is simply to inquire about the outside world in order to determine what is the case there. It might seem that, in determining whether I *believe* that there is an apple tree in my backyard, what I would do is to inquire about my mental world by looking for the very belief. Transparency theorists suggest that the method that we use to determine whether we have a certain belief is essentially the same as the method we use to determine whether the world is such and such. That is, we still focus on what happens in the outside world when making claims that supposedly belong to the mental world. The inspiration for this view comes from the following passage in Gareth Evans:

> [I]n making a self-ascription of belief, one's eyes are, so to speak, or occasionally literally, directed outward upon the world. 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 outward phenomena as I would attend to if I were answering the question "Will there be a third world war?" 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*.[1]

Transparency accounts in general hold that when facing the question "Whether I believe that *p*?" the subject's attention should not be directed inwardly to the state of

[1]Gareth Evans, *The Varieties of Reference*, ed. J. McDowell (Oxford UK: Clarendon Press, 1982), p. 225.

believing but outwardly to the content of the belief.[2] From the first-person point of view, one can settle the question "Whether I believe that *p*" by settling the question "Whether *p*?" or "Is *p* true?" For example, if after considering the reasons for and against the possibility of a third world war, I give an affirmative response to the question "Is there going to be a third world war?," then I am justified in attributing to myself the belief that there is going to be a third world war.

Transparency theorists (almost) universally accept what I call the "transparency claim":

> Transparency Claim: We answer the question of whether I believe that *p* in the same way as we answer the question of whether *p* is true.

However, the theory's proponents disagree about why the transparent claim is true. For example, Alex Byrne[3] holds that the transparency claim contains an inference from the premise "*p*" to the conclusion "I believe that *p*" and that since this kind of inference is self-verifying, it is guaranteed that we can attribute beliefs to ourselves via the transparency claim. Moran, however, does not maintain this inferential view. Since Moran's transparency account is the main focus of this paper, I shall only concern myself with what Moran thinks justifies the transparency claim in making correct self-ascriptions of belief.

2. MORAN'S TRANSPARENCY ACCOUNT

According to Moran, the transparency claim exploits the concept of belief:

> Any understanding of belief that provides for the minimal idea that believing involves "holding true" will entail that it is at least possible to announce one's belief by reporting on the truth as one sees it. If my intention is to report on my belief as such, and I know (how could I fail to know?) that my belief about X is what I hold to be true of X, then my intention will not be thwarted if I make this report by considering what is true of X.[4]

Although it may be too strong to interpret Moran as identifying the concept of belief with the concept of holding true, we can at least accept that for Moran there is a conceptual relation between believing that *p* and holding that *p* to be true:

> [F]rom the first-person point of view, the relation between one's own belief and the fact believed is not evidential or empirical, but is instead categorical. . . . Referring to a *categorical* rather than an empirical relation here is a way of saying that to be a believer at all is to be committed to the truth of various propositions. "Taking my beliefs to be true" is not an empirical assumption of the sort that I might make

[2]In Evan's case, the belief in question is about the external world, so directing one's attention to the content of this belief is, in a sense, directing the attention to the external world.

[3]Alex Byrne, "Knowing What I Want" in *Consciousness and the Self: New Essays*, ed. Jee Loo Liu and John Perry (Cambridge UK: Cambridge Univ. Press, 2012), pp. 165–83.

[4]Richard Moran, *Authority and Estrangement* (New Jersey NJ: Princeton Univ. Press, 2001), p. 105.

> with respect to another person. Rather, it is the categorical idea that whatever is believed is believed as true. Skepticism itself does not avoid this conceptual relation between belief and truth, for the skeptic renounces belief itself and the commitment it entails.[5]

Based on the context of this passage, we can see that what Moran means by "categorical relation" is "conceptual relation"—that is, my belief that *p* and my commitment to the truth of *p* (or my holding *p* to be true) are conceptually related. Further, to say that they are conceptually related is to say that "to be a believer at all is to be committed to the truth of various propositions." That is, we cannot understand "S believes that *p*" without understanding it as "S holds *p* to be true." It is now clear why the question "Whether I believe that p?" can be answered by answering the question "Is *p* true?" Our answer to the latter, if affirmative, represents our commitment to the truth of *p*, which is *sufficient* for our believing that *p*.

Moran uses the term "avowal of a belief" to refer to the self-ascription of a belief via the transparency claim. Avowals have certain characteristics. *First*, they are exclusively first-personal. I can answer the question "Whether I believe that *p*?" by considering the question "Is *p* true?" However, contemplating the latter is insufficient for determining whether my friend George believes that *p*. *Second*, they do not require the subject who performs the self-attribution to consult any evidence regarding her mental state itself, psychological or behavioral. Since the transparency claim requires the subject to confront the content of her mental state, any evidence for or against the existence of the state itself would be superfluous if its content is about the external world. *Third*, if a person uses the transparency claim in making self-attributions of belief, there is a presumption that they are correct. This presumption is a result of the conceptual relation between believing that *p* and being committed to the truth of *p*. A person who gives an affirmative answer to the question "Is *p* true?" is said to be committing herself to the truth of *p*, and committing oneself to the truth of *p* is sufficient for believing that *p*. Traditionally, when a person's self-ascription of a belief has the above features, she is said to have first-person authority, or privileged access, regarding what she believes. So, we can say that a person avowing a belief via the transparency claim has authority regarding her belief.

However, Moran stresses that there is an aspect of first-person authority that is different from the privileged access in the traditional sense. A person has authority regarding her beliefs if she is responsible for them. This aspect is captured by his transparency account, because when considering the question "Is *p* true?" a person has to consider whether she has reason to endorse that *p*. In this sense, her belief that *p* can be regarded as up to her:[6]

[5]Ibid., pp. 77–78.

[6]It does not mean that a person can endorse a belief at will. Although we have a certain degree of freedom with respect to what to believe, we cannot believe a proposition that we do not have any reason to think it is true. We can at most imagine it is true, or hope it is true, but we can never believe it is true.

> The dimension of endorsement is what expresses itself in one aspect of first-person authority, where it concerns the authority of the person to make up his mind, change his mind, endorse some attitude or disavow it.[7]

One might think that the transparency account can only explain how we can know in an authoritative way about our occurrent beliefs and fails to explain how we can have authority regarding our settled beliefs or regarding our beliefs not produced by deliberation. However, with respect to our settled beliefs, Moran can maintain that our authority originates in the deliberations we performed. Regarding our non-deliberative beliefs, e.g., our perceptual beliefs, Moran holds that we can always choose whether to trust our faculties.[8] For example, if we have reason to think we are hallucinating, we can still manage to distrust our perceptions.

3. TRANSPARENT ATTITUDES

A criticism of the transparency account, offered by David Finkelstein,[9] is that it cannot be applied to our self-ascriptions of attitudes generally, e.g., to our desires. I will focus on Finkelstein's argument and then propose a way to make the transparency claim extendable to attitudes in general.

According to Finkelstein, the main difficulty with applying the transparency claim to our other attitudes lies in how to formulate the transparency relation. In the case of beliefs, we have no trouble in accepting that the question of whether I believe that *p* is transparent with the question of whether *p* is true. However, it is a different story when it comes to our other attitudes, such as wanting or desiring. Finkelstein, first of all, offers what he takes to be the only method to make the transparency claim applicable to other attitudes and then criticizes this method.

Finkelstein suggests that we alter the original the transparency claim to be the transparency claim*:

> Transparency Claim*: We answer the question of whether I believe that *p* in the same way as we answer the question of whether I *ought rationally* to believe that *p* (my emphasis).

Finkelstein thinks that "[t]his shift in starting points has the effect of making the transparency claim* straightforwardly extendable to other attitudes."[10] We can substitute any attitude for the term "believe" in the transparency claim*: if I want to know whether I want X, I just need to consider whether I ought rationally to want X. However, there are many situations where an agent has an attitude contrary to what he ought to have. There are two examples in the following passage:

[7]Moran, p. 92.

[8]Richard Moran, "Replies to Heal, Reginster, Wilson, and Lear," *Philosophy and Phenomenological Research* 69 (2004): 455–72.

[9]David H. Finkelstein, "From Transparency to Expressivism" in *Rethinking Epistemology*, vol. 2, ed. Günter Abel and James Conant (Berlin: De Gruyter, 2012), pp. 101–18.

[10]Ibid., p. 104.

> Max has a spider phobia. He is terrified of spiders—spiders in general, and the one on the pillow beside him in particular. Max doesn't take this fear to be rational; he knows that he suffers from a phobia. Still, when he says, "Get that thing away from me; I'm really afraid of it!," he is speaking (or shouting) about his own state of mind with first-person authority. . .. I sometimes look after my friend Adam's dog, Sadie, when Adam is out of town. I don't view it as rationally incumbent upon me to be fond of Sadie. I cannot answer the question of whether I am fond of Sadie by addressing a question about whether I ought rationally to be. Nonetheless, I'm fond of Sadie, and I have no trouble speaking with first-person authority about my fondness for her.[11]

Does Finkelstein propose a reasonable objection against Moran's transparency account? I do not think that Finkelstein captures the spirit of Moran's transparency account. The transparency claim* maintains that the question "Whether I believe that *p*?" is transparent with the question "Whether I ought rationally to believe that *p*?" However, we should note that the relation between "belief that *p*" and "ought to believe that *p*" is not *categorical*. Consider a situation where a person who does not believe that *p* still thinks she ought to believe that *p*. People finding the famous *Pascal's Wager*[12] convincing may be in this kind of situation. They do not believe that God exists, but they recognize that they should. On the other hand, the relation between "believing that *p*" and "holding *p* to be true" is so tight that we cannot understand someone's believing that *p* if we do not understand her as holding *p* to be true. It is this relation that Moran thinks a person is justified in using to self-ascribe beliefs. The transparency claim* contains no such relation, so it should not be regarded as a position that will be accepted by Moran.

We have seen that the adoption of the transparency claim is warranted for the concept of belief: To believe something is to hold that thing to be true. If we want to achieve the same success with other attitudes, it will be instructive to consider their respective concepts. I believe that as long as we can make sense of their concepts, we can apply the transparency claim to them. I will now show how Donald Davidson's view about the nature of attitudes can contribute to this project.

On Davidson's view, to make an agent's action reasonable in light of a certain belief and pro-attitude is to construct the practical reasoning or practical syllogism for that agent.[13] And what Davidson does in constructing the practical syllogism is

[11]Ibid., p. 105

[12]*Pascal's wager is an argument used to show that we should* believe in God. It usually goes as follows: Either God exists or not. (1) If God exists and we do not believe in it, we will go to Hell after the Day of Judgement. However, (2) if we believe in God and He does exist, we can receive salvation and ascend into Heaven. (3) If God does not exist and we believe that He exists, we just have false beliefs. And (4) if God does not exist and we do not believe it, we are lucky to have true beliefs. Now, case (1) is the worst, while case (2) is the best. Case (4) is slightly better than case (3), which is a kind of minor mistake we would make in everyday lives. From this, we can see that not believing in God may lead to (1) if we are wrong but only rewards us with (4) if we are right. Thus, anyone who recognizes the possible consequences *ought rationally* to believe that God exists.

[13]Donald Davidson, "Actions, Reasons, and Causes" in *Essays on Actions and Events,* Donald Davidson (Oxford UK: Clarendon Press, 2001), p. 9. (The article was originally published in 1963.)

to treat beliefs as *cognitive* judgments and pro-attitudes as *evaluative* judgments, that is, my belief that *p* is my judgment that *p* is true, and my desire that *p* is the judgment that *p* is desirable.[14] Since we agree that we cannot understand someone's believing that *p* without understanding it as someone's holding *p* to be true, it is natural for us to embrace the idea that belief is cognitive judgment.

It may, though, seem controversial to treat pro-attitudes as evaluative judgments. After all, a subject can judge that *p* is desirable without wanting that *p*.[15] A familiar case would be a chain-smoker who *judges* that it is desirable to quit smoking because of long term health concerns but does not want to quit. To address this issue, we need only relativize the degree of desire that the smoker attaches to each particular desideratum. Specifically, we should take into account the desirability of smoking in the eyes of the chain-smoker. Hence, the apparent counterexample does not show that a person's desire that *p* is not his judgment that *p* is desirable, but that a judgment does not necessarily lead to a corresponding action all on its own, since there are always additional determinants involved in practical reasoning.

As to the attitudes other than desire, Davidson thinks that "[w]e may suppose different pro attitudes are expressed with other evaluative words in place of 'desirable.'"[16] Following this idea, we can formulate the transparency claim for attitudes as follows:

> Transparency claim[a]: We answer the question of whether I desire, hate or love that *p* in the same way as we answer the question of whether *p* is desirable, abominable or lovable.

Since my judging that *p* is desirable, abominable or lovable is essential to my wanting, hating or loving that *p*. Now, take Finkelstein's fondness for Sadie for example. Rather than asking himself "Whether I ought to be fond of Sadie?" Finkelstein should consider "Whether Sadie is cute, or lovable, or adorable?" If his answer is "Yes!" he is justified in assuming that he has that attitude towards Sadie, for we cannot imagine someone's being fond of *x* without judging *x* to be lovable just like we cannot imagine someone's believing that *p* without judging *p* to be true.

4. TRANSPARENT EMOTIONS

I have shown how it is possible to construct the transparency claim for desire and, supposedly, for attitudes in general, but my method does not immediately eschew all the counterexamples proposed by Finkelstein. His counterexamples involving emotions are, in particular, tricky to deal with, for example, Max's fear of spiders.

[14]Donald Davidson, "Intending" in *Essays on Actions and Events,* Donald Davidson (Oxford UK: Clarendon Press, 2001), p 86. (The article was originally published in 1978.)

[15]This kind of objection can be found in Lauren Ashwell, "Deep, Dark . . . or Transparent? Knowing Our Desires," *Philosophical Studies* 165 (2013): 245–56.

[16]Davidson, "Intending" in *Essays on Actions and Events*, p 86.

However, in "Hume's Cognitive Theory of Pride"[17] Davidson developed an account based on Hume's writings of what causes and rationalizes human pride, which, Davidson thinks, can be applied to emotions in general. In this section, I will show how this account can help us to make the transparency claim extendable to emotions.

Davidson restricts the type of pride he attempts to explain to propositional pride. Propositional pride is pride *that*, which can be contrasted with, for example, pride *of* or *in* something. He thinks that "cases of being proud of something (or taking pride in something, or being proud to do something) reduce to, or are based on, propositional pride."[18] I understand the reduction to imply that every sentence describing a person's non-propositional pride can be rewritten as a sentence about that person's propositional pride without changing what it is that the person takes pride in. So, the sentence "A person is proud of his car" can be rewritten as "A person is proud that he owns the car he does."

Referencing Hume, Davidson accepts that belief and attitude are causally related to pride:

> [T]he cause consists, first, of a belief concerning oneself, that one has a certain trait, and second, of an attitude of approbation or esteem for anyone who has the trait. Together these result in self-approval or self-esteem—what is normally called pride.[19]

Further, Davidson thinks the emotion can be rationalized by its causes in the same way that we rationalize actions:

> The causal relation echoes a logical relation. . . . The causes of pride are a judgment that everyone who exemplifies a certain property is praiseworthy and a belief that one exemplifies that property oneself. The causes of pride are thus judgments that logically imply the judgment that is identical with pride.[20]

The last sentence of this passage suggests that Davidson maintains that emotions are judgments. If emotions are judgments, which can be inferred from the beliefs and judgments that cause them, then it will be simple to construct transparent questions for emotions. Take my pride in owning a beautiful house, for example. I just consider the questions "Is it true that I own a beautiful house?" and "Is it praiseworthy for anyone to own a beautiful house?" If my answers to both are affirmative, I can attribute to myself the pride in owning a beautiful house, given that this pride is logically implied by what causes it. Similarly, in the case of Max's fear that the spider is present in his room, Max just need consider the question "Is it true that there is a spider in my room?" and "Is it dangerous for anyone who has a spider in his room?"; his answers, if affirmative, will produce the judgment and belief that necessarily imply the fear.

[17]Donald Davidson, "Hume's Cognitive Theory of Pride" in *Essays on Actions and Events,* Donald Davidson (Oxford UK: Clarendon Press, 2001), pp. 277–90. (The article was originally published in 1976.)

[18]Ibid., pp. 277–78.

[19]Ibid., p. 284.

[20]Ibid.

Some commentators are reluctant to interpret Davidson as holding the view that an emotion is a judgment.[21] I also have misgivings. Scholars studying emotions have pointed out that they are complex mental states that involve *physiological responses* that are not in the direct control of an agent, while judgments are something that an agent can decide whether to have or not. I agree with this, but it is worth noting that emotions are *intentional* and not identical with physiological responses that are completely severed from our other attitudes. We may find absurd the claim that my *faster breathing* engendered by my seeing the spider is *about* that spider, but we have no trouble accepting the claim that my fear is about that spider. Moreover, emotions are also *rational* in the sense that they can justify or be justified by other attitudes; that is, they engage in the justificatory relation with other attitudes. For example, I am afraid of that spider, *because* I believe that it is dangerous and poisonous. With intentionality and rationality being indispensable elements of emotions, emotions should never be equated with physiological responses.[22]

So, what is the relation between the emotion and the belief and judgment that can be used to rationalize it? We might attempt to understand the relation to be one of *sufficiency*. I have the fear that the spider is present in my room *if* I believe that a spider is present in my room and judge that it is dangerous for anyone who has a spider in his room. We could thus determine whether I fear that the spider is present in my room by considering the transparent questions "Is it true that the spider is present in my room?" and "Is it dangerous for anyone who has a spider in his room?" Unfortunately, it would be incorrect to understand the relation between an emotion and its engendering belief and judgment to be that the latter is sufficient for the former. What Davidson has said, and I agree, is that the relation is causal, and the ordinary view on the causal relation is that the presence of a cause is neither necessary nor sufficient for the effect.

How can the idea that the belief and judgment *typically* cause the emotion help us in responding to Finkelstein's counterexample involving the fear of spiders? Here I propose a weakened transparency account for fear:

> Due to a causal relation, we can answer the question of whether I fear that the spider is in my room by answering the question of whether it is true that the spider is present in my room and whether having a spider in one's room is dangerous?"

This account can be generalized to accommodate other emotions:

> Transparency Claim^{e}: Through a causal relation, we can answer the question of whether I α that *p* by answering the question of whether *p* is true and the question of whether the *p*-type case is β. (α is a certain emotion; β is a feature of the *p*-type case; and the emotion α and the feature β are related in the sense that a person's α that *p* can be justified by the person's belief that the *p*-type case is β.)

[21]See, for example, Mitchell Green, "The Rationality of the Emotions" in *A Companion to Donald Davidson*, ed. Ernest Lepore and Kirk Ludwig (Chichester UK: Wiley-Blackwell, 2013), pp. 506–18.

[22]Such view about emotions can be found in Peter Goldie, "Emotion," *Philosophy Compass* 2/6 (2007): 928–38.

Since the belief and judgement are only causally related to the emotion, they are neither necessary nor sufficient for it. The transparency claim^{e} is only a generally reliable approach to self-knowledge.

However, I think that the transparency claim^{e} is the most we can ask for given the difference between the nature of emotions and that of beliefs and attitudes. According to Davidson, our beliefs and attitudes are our cognitive and evaluative judgments that we can actively form by considering the relevant reasons for and against their contents, while our emotions are only the effects of those judgments we actively form. In this sense, our emotions are not something that can be determined by us. However, this does not mean that emotions completely escape the considerations of transparency questions. Emotions are still causally related to other beliefs and attitudes and can be justified or rationalized by them. Thus, contemplating the transparency questions in the transparency claim^{e} still sheds light on what our current emotion is.

5. FINKELSTEIN'S EXPRESSIVISM

Finkelstein might still be unsatisfied with the transparency claim^{e} and believes that his theory explains self-knowledge of emotions better. To complete my defense of Moran's transparency account, I want to show that Finkelstein's theory fares no better in explaining how we know our emotions. In *Expression and the Inner*,[23] Finkelstein proposes his account of first-person authority, which is consistent with the traditional view, i.e., there is a presumption of truth attached to a person's self-ascriptions of mental states, and that this presumption does not attach to non-self-ascriptions of mental states. Finkelstein points out that the self-ascription of mental states performs a dual function.[24] When an agent verbally expresses what she believes, wants or fears by uttering "I believe that . . . ," "I want that . . . " or "I fear that . . . " the function of her expression is the same as that of her facial expressions which can *manifest* her belief, desire or fear. In this respect, her utterances are the manifestations of those mental states just like her facial expressions are. On the other hand, Finkelstein maintains that these utterances also have an assertoric function: they *state* the condition of a person's mental state and thus can be either true or false. Given this dual function view, it is not difficult to see why there is a presumption of truth attached to self-ascriptions of mental states. If a person utters "I fear that spider" and her utterance is accompanied by a manifestation of that fear, then her assertion will be true, because her manifestation verifies her assertion.

Does Finkelstein propose an adequate account of how we speak our minds with first-person authority? I am afraid he does not. We have to note that Finkelstein implicitly assumes that a person's utterances always manifest the corresponding emotions, and, once we start to consider what prompts a person's utterance, we will find this assumption to be problematic.

[23]David H. Finkelstein, *Expression and the Inner* (Cambridge MA: Harvard Univ, Press, 2003).
[24]Ibid., pp. 101–02.

Let us return to the case of Max's fear of the spider and ask what is it that triggers Max to shout "I am afraid of that spider." There are two candidates: Max's physiological responses to seeing the spider, and Max's beliefs about the spider. If we take those responses to be the trigger, we can imagine a situation where Max's heartbeat accelerated as soon as he caught a glimpse of the spider, and that the accelerated heartbeat prompted him to shout that "I am afraid of that spider"; yet, after a moment of reflection, Max realized that he did not believe the spider to be fearful at all, recognizing it to be his brother's tarantula. In this situation, what Max's utterance expresses is his physiological responses, engendered by the spider, rather than his fear toward that spider, since the fear, as a kind of emotion, must be caused by the beliefs concerning the fearful aspects of that spider (according to Davidson's analysis of emotions). Of course, in other situations, Max's utterance would express his fear if he discovered that he believes that the spider is fearful. Nevertheless, pointing out the former situation is enough to establish that it is not necessary for one's emotion to occur whenever one is prompted to utter a sentence that *appears* to describe that particular emotion. Without this necessity, Finkelstein's account of first-person authority regarding emotions is at most as good as our defense for Moran or even worse, because our transparency claim[e] at least relies on the causal relation between beliefs and attitudes and emotions, while Finkelstein's expressivist account depends only on the chance that one's verbal expressions are prompted by one's beliefs.[25]

[25]I would like to thank Ruey-Yuan Wu for her comments on the earlier draft of this paper. Many ideas came from the discussion with her. Also, I benefitted greatly from the discussion with Chi-Chun Chiu. He pointed out how I could improve my arguments.