

Logique & Analyse 223 (2013), x–x

A PUZZLE ABOUT KNOWLEDGE IN ACTION

ISKRA FILEVA

Introduction

There must be a connection between acting for reasons and knowing what we are doing and why. This is evidenced by the fact that sometimes, we answer questions about reasons for action by saying something about knowledge in action, rather than reasons. Thus, suppose I ask you, “Why are you tapping on the floor?” There are various things you can say in response — for instance, you can say that you are typing and that tapping while you are typing is a habit of yours, or that you are tapping to the beat of a song you are quietly humming to yourself, or that tapping is a good exercise for the ankles, etc. All of these answers cite possible reasons. But you may, instead of giving me some such reason, say simply: “Oh, I did not realize I was doing that.” This latter reply is not telling me *anything* about your reasons for acting. So it may seem that, in one sense, it is not a reply to my question at all. Yet, “I did not realize I was doing it” is a perfectly legitimate response in this case: it would certainly not do for me to follow up on it by insisting, “OK, but I did not ask you whether you *realized* that you were tapping on the floor. What I wanted to know is *why* you were tapping. You must tell me *that*.” I cannot do this because it seems clear that if you were not aware of the fact you were tapping on the floor, you could not possibly have done it *on the basis of reasons*. But it follows from here that there is a connection between acting for reasons and knowing what one is doing and why.

What is this connection?

Many philosophers of action, especially those of a non-causalist persuasion, have built on our intuitions about cases such as the one I just described in order to argue that the connection between knowing what one is doing and acting for reasons is a necessary one in this sense: reason-based action *entails* knowledge of what one is doing. One cannot act for reasons unless one has knowledge of one’s own action. Anscombe (1963), Velleman (1989), Moran (2001), and Wilson (2000), to mention but a few prominent

examples, have all supposed this.¹ The view reflects both our pre-theoretical intuitions about cases such as the tapping one above and a certain very plausible theory of reasons and reason-based action: reasons, on that theory, are considerations which *guide* action and which must be distinguished from blind forces — such as impulses and irrational phobias — that may *push* us around in the absence of clear knowledge of what we are doing and why. But one cannot be *guided* by reasons if one does not know *what* those reasons are. Hence, acting for reasons implies knowledge of *what* one is doing and *why*.

There is, without doubt, something right about this picture — there is an important connection between acting for reasons and knowing what you are doing, and this connection, surely, has something to do with the fact that reasons are not blind causal forces but rational motivating forces. An impulse or a nervous tic which make one tap on the floor without being aware of it do not count as *reasons*, and the action they effect does not count as an action *done for reasons* either. But, the story goes on, reasons are different from nervous tics: reasons move us not by making our bodies react in certain ways but by convincing us to pursue particular courses of action, by presenting considerations we *understand*. So when we act for reasons, we know *why* we are doing what we are doing.

Intuitive, though, as this picture may be, there is a complication: we are sometimes *mistaken* about our reasons for action. But reasons partly determine the description of an action and when we are in error about our reasons for acting, we are in error about what we are doing. So we are sometimes mistaken about what we are really doing. Yet, actions done in ignorance, or partial ignorance, regarding one's own reasons may well count as reason-based. It follows that the connection between acting for reasons and knowing what one is doing is looser than entailment: one can act for reasons and *not know* what one is doing. Consider a preliminary illustration. Suppose Emma's elder sister Jane opens a letter sent to Emma. Imagine she does that slowly and after a period of deliberation. At first she wonders whether she should unseal the envelope, hesitates for a moment, then decides to go ahead and break the seal. Suppose also she acts in the belief she must know the contents of the letter, say in order to be able to protect Emma from someone's dangerous influence. However, she is a curious person by nature and, although a desire to protect Emma is what she *thinks* motivates her, she is *actually* motivated mainly by curiosity. If this description of Jane's motivation is accurate, did Jane act for reasons?

¹ I note in passing that these philosophers differ with regard to the question what *type* of knowledge is the knowledge at stake — is this knowledge non-observational, for instance? Is it non-inferential? Etc. For present purposes, we need not dwell on these differences.

It seems so: for what *else* could Jane's action have been based on if *not* reasons? Her opening Emma's letter was neither instinctive, nor impulsive, nor unintentional. It was preceded by deliberation, there was no coercion or absence of self-control on her part. This is the type of action for which an agent is praised, blamed, and held accountable. It is, for all intents and purposes, an action done for reasons. But Jane is mistaken about *why* she acts and, from here, about *what* she does. For she thinks that all she does is *attempt to protect her younger sister* while, in fact, she *attempts to satisfy her own curiosity*.

Still, there must be *some* connection between acting for reasons and knowing what one is doing and why. If there wasn't, your answer "I did not realize I was doing that" in response to my question "Why are you tapping on the floor?" would not be an answer to my question at all: if there is *no* connection between knowledge of what one does and acting for reasons, I can, on good grounds, follow up on your response by saying, "I am not asking you whether you *realized* that you were tapping on the floor, what I want to know is *why* you were tapping."

Here, we have the seed of a problem which standard causalist belief-desire models of action, on the other hand, face. This problem has been detected earlier by David Velleman and Kieran Setiya, and it is the following: belief-desire models have no resources to explain why there should be *any* connection between acting for reasons and knowing what one does and why. Thus, imagine an agent motivated by a desire and a belief who has *no* knowledge of why he does what he does. For instance, suppose you are tapping on the floor motivated by the desire to interrupt my work, but you are quite unaware of that desire. Indeed, you think you would be glad to help me with my work, for you are a good officemate. And it is true that if I request help from you, if I ask you to copy an article for me, for instance, you would do it (as you have in the past). Still, it may well be that, although you see yourself as a good officemate and this is how I see you, too, you wish, unconsciously, to see me fail, and you are tapping on the floor motivated by the desire to interrupt my work so that I may miss the deadline for my project. Note that on the belief-desire model, your tapping in this latter case qualifies as a reason-based action because it meets the necessary and sufficient conditions on reason-based actions: it is properly motivated by a desire and a belief. Yet, intuitively, it isn't a reason-based action, and there is a problem with a theory which implies that it is.

Setiya suggests that in order to avoid this type of problem, we must incorporate practical knowledge — understood as knowledge of what one is doing and why — as a *necessary* condition of reason-based action. Setiya's proposal, I will argue, has important virtues, but it shares the weakness of standard non-causalist models in that it goes too far with the knowledge requirement: *knowledge* of what one is doing is not a necessary condition of

reason-based action since one can be wrong about what one is doing and why, and still act for reasons; some *connection to knowledge* is.

Now we have a sketch of the “puzzle about knowledge in action” I referred to in the title of this essay: on the one hand, it appears that acting for reasons and knowing what one is doing are importantly connected; on the other hand, an agent can act for reasons but be mistaken about what she is really doing and why. What, then, *is* the relation between knowledge in action and acting for reasons? In order to provide an answer to this question, we need a model of reason-based action which, unlike standard belief-desire models, helps make sense of the connection between acting for reasons and knowing what one is doing, but unlike standard non-causalist models, does not postulate a *necessary* connection between reason-based action and knowledge of what one is doing and why. What I intend to do here is develop such a model.

1. *Setiya on the problem with the belief-desire model and its solution*

I suggested above that standard causalist belief-desire models cast the net too widely: actions such as your tapping on the floor motivated by the unconscious desire to interrupt my work come out as reason-based on those models. And I mentioned that the shortcoming of such models is in that they fail to incorporate a connection to practical knowledge, or knowledge of what one is doing and why, as one of the conditions of reason-based action. But even if belief-desire models are too permissive in the way I claimed, it is not *obvious* that the missing ingredient is a connection to practical knowledge. Perhaps my suggestion was too hasty, perhaps what is missing is something else.

Kieran Setiya (2007) discusses this question, considers two other possibilities, and rejects them in turn. I believe he does so rightly. The problem is that his own positive proposal does not work. Below I will follow him up to the point at which he and I part company.

In making his case, Setiya takes cue from a story told by Freud and retold by Velleman.² The story is autobiographical and goes as follows: one day, Freud’s sister enters his room and comments on the inkstand on Freud’s desk. She says that Freud’s beautiful desk is marred only by the presence of Freud’s old inkstand. Later, Freud sweeps the inkstand in a particularly clumsy way. According to Freud’s own interpretation of the case, in sweeping the inkstand he was motivated by the desire to get his sister buy him a new inkstand and the belief that she will buy him one should the old one get broken. But, as both Velleman and Setiya note, Freud’s action clearly does

² Velleman (2000, pp. 2–3) quoted at Setiya (2007, pp. 33–35).

not qualify as “reason-based.”³ So an action performed on a hidden motive *only* does not count as done “for reasons.”⁴ Setiya calls this type of action “merely motivated”, and I will adopt the same expression, for lack of a better term. The question is, why doesn’t Freud’s action qualify as reason-based? *What* is missing?

The most obvious reply, perhaps, would be — intention: Freud does not *decide* to do what he does, and he does not *intend*, at least not as we normally use the word “intend”, to achieve a particular end. Setiya considers this possibility but rejects it. He does so on the ground that there are cases of performing an action *intentionally* which aren’t, for that matter, cases of *acting for a reason*.⁵ He illustrates this claim with the following sketch: “Suppose I am convinced that I ought to be a lawyer, but only because I am pressured into it by my parents. As I go through law school I truly believe that I am suited to this kind of work. I do not respond to the clues that indicate otherwise: the fact that I spend much less time working than my peers, that I often feel lethargic, that I never get good grades. I would never act on these reasons as grounds on which to quit. Still, I might decide to quit, and be moved to do so, unconsciously, by beliefs that correspond to these facts — finding my own decision both capricious and hard to explain.” (Setiya, 2007, p.35).

Setiya correctly concludes from here that intention alone does not suffice to turn a *merely* motivated action into one done *for reasons*. The more interesting question is, if intention doesn’t make the cut, what does?

A second possible route to take is to suppose that an agent can be said to act for a reason if and only if he *sees* his reason as a *good* one and as a

³ Setiya (2007, p. 33) writes that Freud’s behavior: “meets the conditions of the standard model (i.e. the belief-desire model-my clarification). It is caused by the desire for an end and a belief about how to achieve that end. And this causation is non-deviant, since the desire to break the inkstand guides the movements of Freud’s arm in sweeping it to the floor. Despite all this, Freud did not intentionally break the inkstand and he did not break it for a reason”. It is worth noting, however, that the account which Setiya proceeds to offer is a *causalist* one. In Setiya’s view, an action is caused by intention, and the intention plays a causal role in sustaining the action.

⁴ Note that there have been some attempts to see Freudian style explanations as “based on reasons” and Freud himself as showing that behavior we’d otherwise take to be irrational is really rational. On the former point see Church (2005), Church (1987), and Bermudez (2002). For a qualified defense of the latter, see Alexander (1962). The account I develop here should not be lumped in the same category as those of authors such as Church and Bermudez. While I do believe that the rationality of unconscious and subconscious states is a topic worth exploring, and I do share with Church and Bermudez the view that an action can qualify as reason-based even if we don’t know why we do it, I do not subscribe to the thesis that unconscious mental states *themselves* count as *reasons*. What I say in this essay should suffice to demonstrate why.

⁵ This is a change in Setiya’s view regarding the relation between reasons and intentionality. Cf. Setiya (2004, p. 370).

sufficient justification of his action. The reason may not, in fact, be good: it suffices that the agent *takes* it to be good, at least on that particular occasion. This theory, which has been dubbed the “guise of the good” theory, is traceable back to Plato and Aristotle and has been, more recently, defended by Anscombe (1963), Velleman (1989), and Raz (1999). If the view is tenable, we shall have a satisfactory answer to the question why Freud in the inkstand case does not act for reasons: Freud does not take breaking-the-inkstand-in-order-to-make-his-sister-buy-him-a-new-one to be a *good* reason for action. It seems intuitively plausible that *if* Freud had taken himself to be acting on *good* reasons (for instance, if he had made a conscious decision to break the inkstand in order to make his sister buy him a new one, and had taken that to be a *good* reason for breaking the inkstand), then his action *will* count as done for reasons.

Setiya, however, joins discussants such as Michael Stocker (1979) in the view that it is possible to act for a reason one considers bad. He illustrates the point with an example adapted from Burnyeat (1980): “. . . consider (. . .) someone who enjoys philosophy for the sense of power it can give, even though he does not see such pleasures as worthwhile in the least. He asks derisive questions at talks because that will humiliate the visiting speaker. This is his reason for acting — he does so intentionally — but he recognizes all the while that it is not a good reason to act,” (Setiya, 2007, p.37).

Setiya concludes from this example that one can act for a reason which one does not see as “good” in any way. The “guise of the good” pathway, therefore, leads to a dead end, too.

This objection appears sound to me, as well. But I wish to make two further points in this regard. It is possible to counter the objection by saying that the philosopher in the described example does see something good in his reason for action — the action satisfies his own desire for dominance and is expressive of his lack of empathy. If he had no desire for dominance and did not lack empathy, if witnessing visiting speakers humiliated provoked in him compassion, not pleasure, it would be hard to make sense of his action — the action would become unintelligible: we see the action as intelligible only because we suppose that the philosopher in question finds a dark pleasure in what he does. So the action does present itself to him under the “guise of the good”, although he may not believe that the pleasure he takes in humiliating the visiting speaker is a good reason to act all things considered.

This counterargument would be misguided because the “guise of the good” theory is meant to be *normative*: an agent acting, according to this theory, sees not simply “something good” in what he does (such as the anticipated pleasure of humiliating someone), he takes his reasons to be *good reasons* for action. This normative theory is untenable: while it is true that *in the usual case*, people see their reasons as good (and sometimes, attribute to

themselves reasons they do not actually have in order to avoid the conclusion that they are acting on bad reasons), this is not *always* the case.

But one can, perhaps, propose a revised version of the “guise of the good” theory, one purged of normative overtones. Maybe we can say that an agent acts on reasons when he sees *something* good about what he does, even if he does not believe his reasons to be good all things considered.

Even such a revised conception will not work, though. This is my second point. It is possible for an action to fall short of counting as “reason-based” although the agent performing it does see something good about the reasons he attributes to himself. Suppose, for instance, that a person wants to know what time it is. However, instead of looking at his wrist watch, he goes to check the clock painted on a canvas hanging in his living-room. It may well be that this person *does* see something good about his reason for action: he does what he does because he wants to find out what time it is; but his action does not count as done “for reasons”. I shall return to the question “why not” shortly. The answer will help us specify a condition on reason-based action. Let me now go back to my main line of argument.

If neither intention nor the “guise of the good” is the missing ingredient here, what is? Setiya suggests the following: what distinguishes actions with purely psychological explanations, such as those of Freud or the willy-nilly-about-to-become-a-lawyer person, from actions *done for reasons* is that when we act for reasons we know *what* we are doing and *why*. Those philosophers of action who subscribe to a normative view of reason-based action are quite right to maintain that action-done-for-reasons involves beliefs about oneself and one’s reasons and is thus *reflective*; but “reflection” here, he insists, should be understood as the activity of *recognizing* one’s reasons as one’s own, not of *evaluating* them, as *reflexivity*, rather than *reflectivity*. He says: “And in writing about the “reflective” character of acting for reasons, I have in mind not reflection as an activity of its own, but the fact that a belief about oneself is involved in taking something as one’s reason to act. The point is simply that in acting for a reason, I *know* (my italics) what my reason is (. . .). While I may not have thought about it consciously, I do not need to investigate myself, as I would another person, in order to give my reason for acting: I already know,” (Setiya, 2007, p.47). And earlier: “I don’t have to ask myself why I am walking to the shops if my reason for doing so is to buy a hammer,” (Setiya, p.40).

His arguments for this conclusion are nuanced and subtle, and I cannot do full justice to them here but, for present purposes, this much will do: despite the elaborate reasoning behind it, this conclusion doesn’t work. It doesn’t because it cannot help explain why Jane’s action in the case I began with counts as one done for reasons although Jane is wrong about what she is doing and why. We need, therefore, another model. But before I turn to the task of providing one, I wish to make one more point.

In fairness, in the later of the two pieces of Setiya's which I draw on here, Setiya acknowledges the possibility I am interested in but says that such cases count as at best "marginal" cases of acting for a reason and also that we are: "...properly unsure about the description of self-deceptive action. But it does seem possible," (Setiya, p.44). Since Setiya sees such cases as "marginal", he does not make an attempt to accommodate them. Can we dismiss these cases in the way Setiya does?

The remark I just quoted is cryptic, and I am not sure what exactly Setiya means by "marginal" here — if the thought is that an action done in error of why one acts does not meet the standards of *perfectly* reason-based action, then I agree, and will come back to this point at the end; but so are actions based on *bad* reasons, and Setiya does not count *those* as "marginal." If, on the other hand, he wishes to claim that such actions are very rare, then I disagree. Things may be relatively straightforward with actions such as going to the store to buy a hammer, but they are not at all straightforward with actions such hiring or firing someone (you could hire a minority candidate in the belief this is the most qualified candidate; but you might be motivated, in part, by the desire to prove to yourself you are not a racist, for instance), beginning a new business enterprise (you may think you enjoy the thrill of the risk but, in fact, be motivated by the desire to show to your father you can do this), etc. And even with an action such as going to the store, motivation may not be what it, at first, appears: it might be that I am going to the store to buy a hammer, just as I think I am and that's that; but it might also be that in addition to this, I am motivated to go to the store and buy a hammer *right now* because I want to take some time off my work, although I may well be unaware of this additional motive.

We have now rejected belief-desire models, belief-desire+intention models, reflective endorsement of one's own reasons models (or "guise of the good" models), and Setiya's "reflexive recognition" account. What other option is there?

2. *A proposal*

Setiya has taken a step in the right direction — in accounting for reason-based actions, we need a condition which helps capture the connection between acting for reasons and knowing what one does and why, and Setiya attempts to include precisely such a condition. Setiya is also right in relaxing the constraints postulated in normative accounts: normative accounts require *reflective endorsement* of one's own reasons as a necessary condition of reason-based action. Setiya's model demands merely a *reflexive recognition* of one's reasons. But we need to loosen the constraints further still — even

reflexive recognition is a too restrictive condition. Something weaker suffices: all that is necessary is that the agent possess *some explanation* of what she does and why, that there is something which she takes as her reason. The explanation need not be *true*: the agent need not, properly speaking, “recognize” her reason for what it is — she may *misrecognize* it. The key here is that an agent acting for reasons is able to offer *some* answer to the question: “Why are you/am I doing this?” where the answer cites reasons or putative reasons (and not coercive forces, natural forces, or some such thing); but the answer need not be *correct*. On this proposal, what makes Freud’s sweeping of the inkstand fall short of a reason-based action is the fact that Freud has *no explanation* of what he does and why (it takes post-factum analysis on his part to come up with an interpretation). If Freud had an explanation, some explanation, his action would have qualified as reason-based. For instance, if Freud had, alternatively, broken the inkstand in the belief he wants to check whether the inkstand is sturdy enough to fall on the floor without breaking, while truly motivated by the desire to break it and make his sister buy him a new one, his action would have qualified as based on reasons.

But we need to refine this initial account. In order to see why, let us go back to Jane and Emma. Suppose Jane opens Emma’s letter not in the belief she wants to protect Emma but in the belief she wants to see the weather forecast. In this case, there *is* something which Jane cites as her reason for action, but her action does not count as done for reasons. Why not? The problem is that Jane doesn’t really have an explanation, not even a false one, of what she does; all she has is the *semblance* of an explanation. There is absolutely no ground for her to suppose that the weather forecast could be found in Emma’s private letter,⁶ and so the cited “reason” does not help make sense of her action. Acting for reasons requires an explanation such that the reasons cited therein succeed to make the action *intelligible*.

We can now see why even a revised “guise of the good” account fails: the condition we need is not that one see the reasons one has (or believes one has) “under the guise of the good”, but rather, that those reasons succeed in making one’s action intelligible from an interpreter’s point of view.

There is a long tradition of taking intelligibility to be a mark of reason-based action. Nagel (1970, pp. 33–4.) famously appealed to intelligibility in order to distinguish motivation by reasons from causal deviance. Raz (1999), Dancy (2000), and McDowell (1998) have made similar arguments. My last modification is in line with this tradition, but there is a twist. Other philosophers require that the agent acting be *motivated* by the reasons which make her action intelligible. My account does not require that: the agent

⁶Barring contrived scenarios, for instance, one in which the government keeps the forecast secret from citizens, Emma has an “informant” who mails her the forecast in private letters, and Jane knows this.

acting must possess (or be able to provide) a reasons explanation and one which succeeds to make her action intelligible; he or she need not, however, be actually motivated by the reasons cited in the explanation in question.

But it is important to note that the intelligible explanation which the agent possesses, even if not correct is *not*, thereby, *motivationally inert*. It has positive motivational force in the following sense: a rational agent cares about whether she has an explanation of her actions or not, and would, in general, be deterred from acting if she has no intelligible story to tell to herself about what she does and why. An agent who acts for reasons is partly motivated by the (true) belief that she can make sense of her own actions, even though she may not have a true belief as to why she acts.⁷ What I claim is that this is enough: in order for an action to qualify as reason-based, it does not have to be the case that, in addition to this, the intelligible explanation which the agent possesses is, in fact, *true*, and that she is *really* motivated by the reasons cited in it. It suffices that she is motivated by the fact she *has* an intelligible explanation. Herein lies the truth about a point made by Velleman time and again, namely, that rational agents strive to achieve self-understanding. What *is* true is that rational agents strive for, what one might call, self-intelligibility — we are inclined to abstain from actions for which we can give no reasons to ourselves. It is a further question whether we are interested in genuine self-understanding and so inclined to try to uncover our own hidden motives. Sometimes we are and sometimes we are not. But our actions may qualify as “done for reasons” even when we are not.

There is more truth both to Velleman’s point concerning the drive towards self-understanding and to accounts which postulate a necessary connection between acting on reasons and knowing what one is doing, however, and that truth will point us in the direction of a further refinement of the present account: our reasons for action are *truly* known to us when our actions are described *at a relatively low level of action identification*⁸ such as that of

⁷ If she likes the explanation, she may also find herself motivated, in addition, by the *kind* of explanation which she has (for instance, an explanation which depicts her as a caring and generous person), but this sort of motivation is not, in general, a necessary condition of reason-based action. An action may count as reason-based even if the agent positively dislikes her own explanation of her action (for instance, a Jane who is aware of her real reason for opening Emma’s letter). This is why the normative version of the “guise of the good” account fails.

⁸ Though not the lowest level, since our actions are not known to us at the level of micro-physical processes. I note in passing that Anscombe demurs at the idea of pushing what is known as the content of one’s intention “back” in this way. She writes, “it is an error to try to push what is known by being the content of one’s intention back and back; first to the bodily movement, then perhaps to the contraction of the muscles, then to the attempt to do the thing, which comes right at the beginning.” Anscombe (1963, §30). Here she is motivated by a thought with which I entirely agree: we do not normally think of our actions at the level of bodily movements and may even be unable to perform an action if we tried to think of it at that level — say, I won’t be able to play the Chopin waltz I want to play if I think about the

bodily movements. Jane isn't wrong about why she makes the bodily movements we'd refer to as "opening a letter" — she makes them because she wants to read the letter. What she is wrong about is the reason she has to want to see the letter. If Jane makes the bodily movements we'd refer to as "opening a letter" yet sincerely believes that she does them in order to make a phone call, we would not think she is acting for reasons.⁹ More generally, a person acting for reasons must have *knowledge* of what she does at a relatively low level of action identification.

But we do not think of our actions only at the level of bodily movements. We also give more encompassing descriptions of them. Moreover, we *typically* think of our actions at a higher level of identification, and our reasoning about what to do is, in the usual case, reasoning about what to do *so described*;¹⁰ and we could be mistaken both about why we act and about what we are doing when the action is described at this higher, more encompassing level, although we are correct about what we are doing and why when the action is described at a lower action identification level.

Conclusion

In the beginning of the present paper, I formulated a puzzle about knowledge in action and promised a solution to that puzzle. The puzzle was this: acting for reasons and knowing what one is doing must be importantly connected. If they are not, it would be impossible to answer a question about reasons for acting by saying something about knowledge in action. But we do, quite legitimately, answer questions about reasons by saying something about knowledge, or lack of it, instead. On the other hand, an agent may act for reasons but be mistaken about what she is really doing and why. What then, *is* the precise relation between knowledge in action and acting for reasons? I said we need an account of reason-based action which accommodates actions containing an error regarding one's own motivation and so helps solve the puzzle about knowledge in action. I have now accomplished my task and wish to summarize the results and make a few final remarks.

movements of each of my fingers. What is *not* true is that our intentions at a high level of action identification have to be *known* to us: all that's necessary is that we have an intelligible story to tell to ourselves as to what they are and are partly motivated by the fact we have such a story.

⁹ Again, barring contrived scenarios, for instance, one in which the letter is equipped with sensors, and when Jane opens it, someone's phone rings.

¹⁰ For empirical evidence that we typically think of our actions at the highest level of identification possible, see Kozak, M., Marsh, A.A., & Wegner, D.M. (2008). They present results which support the thesis that we think of our action at low identification levels only when we are incompetent at what we do, or else when we are uncertain about whether we should do something or not.

I argued that a piece of behavior qualifies as an action done for reasons when (a) the agent acting possesses a (possibly putative) reasons explanation of the action; (b) the explanation successfully makes the action intelligible; (c) the agent is at least minimally motivated by the fact she possesses a reasons explanation in the following sense: she would be deterred from acting if she did not possess an explanation of what she does; (d) the agent possesses knowledge of what she does when the action is described at a relatively low level of identification such as that of bodily movements.

Now I wish to qualify the conditions listed further still, and to consider two possible objections to my account. The first qualification is the following. A reason-based action is, if I may put the matter this way, sufficiently *internally* motivated. That means: we can be wrong about our reasons for action but eventual mistakes have to be, in some intuitive sense, *our own*. I am not going to fully specify this sense of “ownership” here, but I will illustrate the point with an example. If an experimenter psychologist feeds my brain the belief that I am performing an action for a reason *a* while in fact, my action is a direct result of the brain states produced in me by the experimenter, my action doesn’t count as done for reasons even though it may meet *all* of the above listed conditions: I do have an explanation of what I do and why, the explanation succeeds to make my action intelligible, and I have knowledge of my action when the action is described at a relatively low level of action identification. While a mistake regarding my own reasons for action is compatible with my acting for reasons, a mistake which comes as a result of a direct intervention in my brain is not so.¹¹

The second qualification is this. I mentioned that the action explanation which an agent possesses must successfully make the action in question intelligible. What I wish to note here is that there is no sharp boundary between explanations which do and such that do not successfully do this. Gray area cases will likely be observed in situations where people act under stress or find themselves in very unusual circumstances.¹²

Turn we now to the objections. I began this analysis on the premise that we can be wrong about our own reasons for action. But why should one

¹¹ It is difficult to say whether this condition rules out actions under post-hypnotic suggestion. I am inclined to say that it does, but I won’t go into this issue here.

¹² Consider an illustrative example. Death mates in Nazi concentration camps are reported to have played the “pretense game” of being Nazis. They made Nazi uniforms and armbands from scraps and gave to each other the salute “Heil Hitler” salute. Surely, they must have had some explanation of what they were doing and why. What explanation could that have been? Suppose they thought that they wanted to have some fun. Does this explanation successfully make the action intelligible in the required way? Should they have been deterred from acting for lack of a better explanation? The answer is, I believe — both “yes” and “no”. The explanation does, indeed, verge on the rationally unacceptable yet, in some sense, it does “make sense” and does confer some intelligibility on the behavior in question. See Bettelheim (1948).

accept such a premise? It is certainly possible to hold that we *always* know our own reasons for action or that we don't have sufficient grounds to assume the contrary. After all, there is, at least, some evidence that the reason which the agent cites to herself has something to do with her action — the evidence is that she cites the reason to herself. But the hidden motive? What can possibly serve as evidence that she has any such? Novelists and psychologists often talk about hidden motives and intentions, but why believe them? Can there be a proof that Jane does not open the letter in order to protect her sister, just as she thinks?

Though I believe that skepticism of this sort is likely to be “in theory” only, the requirement for presenting evidence is quite legitimate, and so the question must be addressed. The nature of hidden motives and intentions is a topic of another and lengthy discussion, but I wish to say something briefly by way of demonstrating the possibility of the phenomenon in question.

The content of our reasons for action is partly determined by what we come to perceive as success and failure in action. Upon occasion, our own perceptions of success and failure may give us a clue as to our “under the surface” motives and goals. Thus, suppose that the letter Jane opens is signed by K. Edwards. Jane believes that “K. Edwards” stands for “Keira Edwards”, a friend of Emma's (one who has, in Jane's view, supplanted Jane in Emma's life and is currently exerting “dangerous influence”). However, the letter is, actually, from Katherine Edwards, a business partner. Upon seeing that the letter is a business letter, Jane feels slightly disappointed — Emma's business is not what she wanted to learn about. Her disappointment might be a signal for her, if she reflects on it, that her real motivation in opening Emma's letter was to satisfy her own curiosity: after all, if *all* she really wanted was to protect her little sister, she should feel glad the letter is from a business partner and not from someone whose influence on her sister is, in her view, dangerous. But she does not feel glad. She feels disappointed. Her disappointment will make sense only if we supposed that the desire which moved her to act was not the desire she thought she had. So much for the existence of hidden intentions.

The second and final objection I am going to consider is the following. The conditions I offer, it can be claimed, are not conditions of *perfectly* reason-based action. One may well have the residual intuition that an action only *really* counts as done for reasons when the agent performing it knows why he acts and is truly motivated by the reasons she gives to herself.

I have no particular quibble with this intuition, actually. I am engaged in, as one might call it, *non-ideal* theorizing here. It is all right to spell out the standards of *ideally* reason-based action, and it is certainly interesting to ponder those as well. But we need to make sense of our ordinary ways of explaining each other's behavior. Just why are we inclined to regard the opening of the letter case as an example of action done for reasons, and

we are not inclined to regard this way Freud's inkstand case or Setiya's law student case? Standards of ideally reason-based action cannot help answer this question.

Our reasons for action are partly transparent to us insofar as our own processes of thinking are. But motives which often remain out of sight influence our actions as well and play a role in the way we make choices. Such motives often require a careful and honest self-scrutiny to be uncovered. What I've argued here is that we need not engage in such self-scrutiny in order for our actions to count as done for reasons; though, of course, *perfectly* reason-based actions are not actions which merely *count* as done on reasons.

University of North Carolina – Chapel Hill
USA
E-mail: ifileva@email.unc.edu

REFERENCES

- Alexander, P. (1962). Rational behaviour and psychoanalytic explanation. *Mind*, 71 (283), 326–341.
- Auscombe, G.E.M. (1963). *Intention*. Oxford: Basil Blackwell.
- Audi, R. (1997). Acting for reasons. In Al. Mele (Ed.), *The philosophy of action* (pp. 75–105). Oxford: Oxford University Press.
- Bermudez, J.L. (2002). Rationality and psychological explanation without language. In J.L. Bermudez & A. Millar (Eds.), *Reason and nature: Essays in the theory of rationality* (pp. 233–264). Oxford, UK: Clarendon Press.
- Bettelheim, B. (1943). Individual and mass behavior in extreme situations. *Journal of Abnormal and Social Psychology*, 38, 417–452.
- Burnyeat, M. (1980). Aristotle on learning to be good. In A. Rorty (Ed.), *Essays on Aristotle's ethics* (pp. 69–92). Berkeley, CA: University of California Press.
- Church, J. (1987). Reasonable irrationality. *Mind*, 96 (383), 354–366.
- Church, J. (2005). Reasons of which reason knows not. *Philosophy, Psychiatry and Psychology*, 12, 31–41.
- Dancy, J. (2000). *Practical reality*. Oxford: Oxford University Press.
- Kozak, M., Marsh, A.A. & Wegner, D.M. (2006). What do I think you're doing? Action identification and mind attribution. *Journal of Personality and Social Psychology*, 90 (4), 543–555.
- McDowell, J. (1998). Functionalism and anomalous monism. In McDowell, J., *Mind, value and reality* (pp. 325–341). Cambridge, MA: Harvard University Press.

- Moran, R. (2001). *Authority and estrangement: An essay on self-knowledge*. Princeton, NJ: Princeton University Press.
- Nagel, T. (1974). *The possibility of altruism*. Princeton, NJ: Princeton University Press.
- Raz, J. (1999). *Engaging reason*. Oxford: Oxford University Press.
- Searle, J. (2001). *Rationality in action*. Cambridge: Cambridge University Press.
- Setiya, K. (2004). Explaining action. *Philosophical Review*, 112, 339–393.
- Setiya, K. (2007). *Reasons without rationalism*. Princeton, NJ: Princeton University Press.
- Stocker, M. (1979). Desiring the bad. *Journal of Philosophy*, 76, 738–753.
- Velleman, J.D. (1989). *Practical reflection*. Princeton, NJ: Princeton University Press.
- Velleman, J.D. (2000). *The possibility of practical reason*. Oxford: Oxford University Press.
- Wallace, R.J. (1999). Three conceptions of rational agency. *Ethical Theory and Moral Practice*, 2, 217–242.
- Wilson, G. (2000). *The intentionality of human action*. Stanford, CA: Stanford University Press.