

# Oxford Handbooks Online

## Making the Unconscious Conscious

David H. Finkelstein

The Oxford Handbook of Philosophy and Psychoanalysis

*Edited by Richard G.T. Gipps and Michael Lacewing*

Print Publication Date: Feb 2019 Subject: Philosophy, Philosophy of Mind

Online Publication Date: Nov 2018 DOI: 10.1093/oxfordhb/9780198789703.013.18

### Abstract and Keywords

How should we understand the distinction between conscious states of mind and unconscious ones? This chapter briefly reviews an answer to this question that the author has set out and defended in earlier work; it then suggests a new answer—one that supplements, rather than replaces, the old answer. In spelling out this new answer, the chapter offers an account of a distinction that is related to, but not identical with, that between conscious and unconscious states of mind, viz. the distinction between conscious and unconscious expressions.

Keywords: expression, expressivism, conscious, consciousness, first-person authority, self-knowledge, unconscious

---

David H. Finkelstein

## 1. A Question

HOW should we understand the distinction between conscious states of mind and unconscious ones? This question, which I'll refer to as 'Q' in what follows, could also be put this way: by virtue of what is someone's anger, fear, anxiety, or desire, for example, rightly characterized as either conscious or unconscious? In what follows, I'll introduce some background to, and then briefly review, an answer to Q that I've set out and defended in earlier work.<sup>1</sup> I'll raise a worry about how this answer could be correct, and I'll discuss an exchange between Jonathan Lear and Richard Moran that bears on how the answer I've given to Q might help to elucidate what Lear calls 'the psychoanalytic meaning of making the unconscious conscious' (Lear 2011: 52).<sup>2</sup> Both the worry and the exchange between Lear and Moran will lead me in the direction of a new answer to Q—i.e. toward another way of understanding the difference between conscious and unconscious states of mind—one that supplements, rather than replaces or corrects, the old answer. Providing it will require that I present an account of a distinction that is

## Making the Unconscious Conscious

---

related to, but not identical with, that between conscious and unconscious states of mind, viz. the distinction between (what I'll call) conscious and unconscious expressions.

(p. 332) A good point of departure for theorizing about what it means to characterize someone's state of mind as either 'conscious' or 'unconscious' is a simple, tempting, and ultimately unsatisfactory answer to Q that can be put as follows: 'A person's mental state is conscious if she is aware of it, i.e. if she knows that she is in it. It is unconscious if she does not know that she is in it'. To see why this is not a satisfactory answer to Q, consider the following story:

Max is a philosophy professor who lives and works in Chicago. Every few months, his mother flies to Chicago from her home in New York to meet with business associates and to visit her only child. Almost every time she does this, Max forgets to pick her up at the airport. Now, as a rule, Max doesn't tend toward this sort of forgetfulness. He never forgets to retrieve his wife or his friends from the airport. Moreover, on those occasions when he forgets to retrieve his mother, he also manages to be away from his mobile phone or to be carrying a phone whose battery has died—so, she is unable to reach him. One day, while talking with a colleague about his mother, it occurs to Max that this behaviour of his might constitute evidence that he is unconsciously angry at her. Indeed, on thinking through some of his other recent behaviour involving his mother, Max becomes convinced by this hypothesis—convinced on the basis of behavioural evidence that he harbours unconscious anger toward her. He says to his colleague, 'I must be unconsciously angry. It's the only way to explain the way I've been acting; I'm angry at my mother'.

Imagine that Max is justified in drawing the conclusion that he does. In such a case, we might describe him as 'knowing that he is unconsciously angry at his mother'. I take it that there is nothing incoherent in such a description. But if this is so, it means that being consciously angry is not the same as knowing that one is angry. At the end of the story, Max knows that he is angry at his mother; he is aware of his anger, but this anger is still unconscious.

Why is it tempting to think that a conscious mental state is just one that its subject knows about or is aware of? I suspect that part of the answer lies in the fact that we use the expression 'conscious of' (nearly) interchangeably with 'aware of'. Right now, as I write these words, I'm aware of—i.e. conscious of—my dog's left front paw, which is resting on my right foot. Of course, no one imagines that it somehow follows from the fact that I'm conscious of Kopi's paw that her paw can itself be described as 'conscious'. But the difference between 'conscious of x' and 'x's being conscious' is easier to get confused about when x is one's own state of mind. In order to get ourselves into a position from which we can think clearly about question Q, we need to distinguish between two ways in which the word 'conscious' is used. At the conclusion of Max's story, he has become aware of—we could say 'conscious of'—his anger toward his mother. But his anger is not conscious. He is conscious of his anger, but he is not consciously angry. (It is only when

## Making the Unconscious Conscious

---

someone is consciously angry—and not merely conscious of his own anger—that his anger is said to *be* conscious.) Q asks how we should understand the difference between conscious states of mind and unconscious ones (i.e. the difference between someone's being consciously, e.g. angry and his being unconsciously angry). We'll fail to get this (p. 333) difference into focus if we conflate it with a distinction between things that one is aware of and things that one is unaware of.<sup>3</sup>

## 2. Wittgenstein's Suggestion

The answer to Q that I've defended in the past may be understood as an attempt to develop a suggestion that Ludwig Wittgenstein makes concerning how to think about psychological self-ascriptions. In his late writings, Wittgenstein often suggests that we should understand such ascriptions as, or as akin to, *expressions*. Thus he writes:

The statement "I am expecting a bang at any moment" is an *expression* of expectation.

(Wittgenstein 1967: §53)

When someone says "I hope he'll come"—is this a *report* about his state of mind, or an *expression*<sup>4</sup> of his hope?—I can, for example, say it to myself. And surely I am not giving myself a report.

(Wittgenstein 1953: §585)

For even when I myself say "I was a little irritated about him"—how do I know how to apply these words so precisely? Is it really so clear? Well, they are simply an expression.

(Wittgenstein 1992: 70)

Wittgenstein thinks that, at least when we're doing philosophy, we are inclined to imagine that saying, 'I expect an explosion', is more like saying, 'He expects an explosion', than in fact it is. Thus, we assume that just as my ascribing an expectation of an explosion to another person requires that I have some epistemic ground or justification for believing that he expects an explosion, so, too, my ascribing an expectation of an explosion to *myself* requires that I have some epistemic ground or justification for believing that I expect an explosion. The likely result will be a view according to which, even under the best of circumstances (even when there is no question of my being opaque to myself), psychological self-ascriptions should be understood as reports of observations—a view that Wittgenstein takes to be misguided.<sup>5</sup> His suggestion—that we think of such ascriptions as expressions—is meant to provide an alternative way of understanding their grammar. When Wittgenstein calls, e.g. a self-ascription of expectation an 'expression', part of what he means is that it is not the report of an observation. Thus, there is a

## Making the Unconscious Conscious

---

fundamental grammatical asymmetry between psychological self-ascriptions and ascriptions of psychological states and goings-on to others.

(p. 334) In Wittgenstein's *Philosophical Investigations*, this suggestion first appears in §244, in the context of an exchange between Wittgenstein and his interlocutor concerning how it is that children manage to learn the names of sensations. During this exchange, we find Wittgenstein saying:

Here is one possibility: words are connected with the primitive, the natural, expressions of the sensation and used in their place. A child has hurt himself and he cries; and then adults talk to him and teach him exclamations and, later, sentences. They teach the child new pain-behaviour.

(Wittgenstein 1953: §244)

It's no accident that it is in the context of a discussion of language learning that Wittgenstein introduces the idea that some self-ascriptions might be understood in terms of expression. Part of his point could be put as follows (and here I'm reading §244 through the lens of many other passages in which Wittgenstein likens self-ascriptions to expressions). No one imagines that when an infant expresses a pain by crying, wincing, or moaning, he is issuing a report concerning an item that he has discovered via some sort of inwardly directed observation. We can think of a child's gaining the ability to say, 'Ouch!' or, 'That hurts!' or, 'I feel a pain in my foot' as a matter of his learning new ways to do something that, in some sense, he could already do before he could talk, viz. express his pains. And while an expression that takes the form of a prelinguistic cry or wince is different in significant ways from an expression that takes the form of a self-ascription<sup>6</sup>—even so, Wittgenstein thinks, there is no good reason to hold that when a child becomes able to produce the latter sort of expression, he suddenly needs to rely on inner observation in order to manage it.

Immediately following the bit of *Investigations* §244 that I quoted in the preceding paragraph, the text continues as follows:

"So you are saying that the word 'pain' really means crying?"—On the contrary: the verbal expression [der Wortausdruck] of pain replaces crying and does not describe it.

The question that Wittgenstein's interlocutor asks in the first of these two sentences is liable to seem oddly confused or unmotivated. After all, Wittgenstein has not suggested (in §244 or elsewhere) that the word 'pain' means something different from what we always took it to mean; he's not suggested that it really refers to behaviour. What is moving the interlocutor to ask what he asks here?

Wittgenstein's interlocutor is in the grip of a false dilemma concerning sensations and other mental items. He (the interlocutor) consistently thinks—not only in the *Investigations* but in other late writings as well—that one must choose between (1) thinking of pains, etc. as things that can be known about only via a kind of inner

(p. 335)

## Making the Unconscious Conscious

---

observation directed at a realm of objects that just one subject can, in principle, have direct access to, and (2) embracing a kind of behaviourism that, in effect, *eliminates* sensations and replaces them with behaviour. In the quoted passage from §244, the interlocutor gathers that Wittgenstein means to be rejecting the idea that we ordinarily self-ascribe pains on the basis of inner observation, so he, in effect, accuses him of embracing behaviourism. In this instance, the interlocutor's way of accusing Wittgenstein of embracing behaviourism is to suggest that Wittgenstein thinks the word 'pain' doesn't after all pick out anything genuinely inner but instead just means crying.

This isn't the only moment in the *Investigations* when Wittgenstein's interlocutor accuses him of behaviourism. But in §244, Wittgenstein replies to the accusation in a way that is liable to sound off-key. He doesn't say merely that an utterance such as 'I'm in pain' should be understood as a verbal expression of pain and not as a description of crying (or of any other behaviour). He says that the verbal expression of pain *replaces* ('ersetzt') crying. And perhaps this doesn't seem quite right. After all, it's not as if children *stop* expressing pains by crying when they learn how to express them by talking about them.

Maybe Wittgenstein ought to have used a different word at the end of §244. Still, I think it's important that there is, after all, a *kind* of replacement that occurs when children learn to express their pains by self-ascribing them. And I think Wittgenstein *might* have had this in mind when he imagined the exchange in the way that he did. In any case, I want to say that there is an important sense in which the *kind* of crying a child did as a prelinguistic infant really is no longer a part of her life after she learns to talk about her own sensations, emotions, and attitudes. The kind of crying she did as an infant is, at that point, replaced by another kind of crying. After she learns to talk, even though her expressions-of-pain-via-crying sometimes look and sound the same as they did before, they now could be said to have a different *form*. I'm going to have to leave this suggestion dark for a little while—until after I've discussed the possibility of an expression's being conscious. I'll come back to it in §7.

### 3. An Answer to Q

In my own writing, I've tried to take what Wittgenstein says concerning the grammar of our talk about our own mental states and use it to make sense of the distinction between two kinds of mental state. The guiding idea could be put this way: once a human being has acquired language, her mental states are such that she is, ordinarily, able to express them via the sort of speech act that Wittgenstein talks about; she can express her sadness, anger, or desire by ascribing it to herself. Insofar as she is able to do this, her sadness, anger, or desire is said to be *conscious*. But sometimes, even after a human being has acquired language (and with it, the capacity to say what she is thinking or feeling), she is unable to express, e.g. some particular fear of hers by self-ascribing it. A kind of (p. 336) expressive ability that is characteristic of normal psychological states and events, as they figure in the life of a linguistic animal, is, in such a case, absent or blocked. The word that we use to characterize this sort of absence is 'unconscious' or

## Making the Unconscious Conscious

---

'unconsciously'; we say that someone's fear is unconscious or that she is unconsciously afraid.

What would it take for Max's unconscious anger toward his mother to *become* conscious? It would not be sufficient for Max to become aware on the basis of behavioural evidence that he was angry. For Max's anger to become conscious, he would need to acquire, or to regain, an ability to express it merely by self-ascribing it. This expressive ability is what he still lacks at the conclusion of the story that I told about him in §1. At the end of that story, Max says to a colleague, 'It's the only way to explain the way I've been acting; I'm angry at my mother'. In so saying, Max expresses a belief (or perhaps knowledge) that he has about himself. But he does not express his anger at his mother. He may express that anger by stranding his mother at the airport, but he is still unable to express it by self-ascribing it. Hence, it is still unconscious.<sup>7</sup> This is to say, I've answered Q by claiming that what distinguishes conscious states of mind and unconscious ones is the presence or absence of a particular expressive ability.

In my *Expression and the Inner*, I put the point as follows:

*Someone's mental state is conscious if he has an ability to express it merely by self-ascribing it. If he lacks such an ability with respect to one of his mental states, it is unconscious.*

(Finkelstein 2003: 120)

What work is the word 'merely' doing in this encapsulation of an answer to Q? Why not leave out 'merely' and say, more simply, that a person's mental state is conscious if and only if he can express it in a self-ascription? The trouble with the simpler formulation is that it is open to counterexamples of the following sort. Imagine that Max occasionally expresses his anger by speaking in a peculiar clipped tone of voice. *While speaking in this tone of voice*, he says: 'The only way to explain my odd behaviour is to posit that I am unconsciously angry at my mother. So, I'm angry at her'. Through his tone of voice, Max expresses his anger in a self-ascription of it. Even so, his anger is unconscious. In *Expression and the Inner*, I wrote:

When I say that someone's state of mind is conscious if he has an ability to express it *merely* by ascribing it to himself, I mean this: the sort of ability at issue is one that enables a person to express his state of mind in a self-ascription of it, where what matters [to the self-ascription's expressing what it does] isn't his tone of voice (p. 337) (or whether he is tapping his foot, or what he is wearing, or to whom he happens to be speaking), but simply the fact that he is giving voice to his sincere judgment about his own state of mind.<sup>8</sup> That I might manage to express my unconscious anger in a self-ascription of it *via a clipped tone of voice* doesn't show that I have the relevant sort of expressive ability. When I am consciously angry, I can say in a neutral tone of voice, "I'm angry," and thereby express my anger.

(Finkelstein 2003: 120)

Was that an adequate response to the tone-of-voice case? In one sense, I think it was. It showed that the case constitutes a merely apparent, not a genuine, counterexample to the view I was defending in the book (and that I still take to be correct). On the other hand, it seems to me now that what I wrote there (and elsewhere) is liable to leave a reader with a nagging question that might be put as follows: ‘Why should it *matter* if a person is able to express his anger merely by self-ascribing it—as opposed to being able to express it in a self-ascription that comes in a clipped tone of voice? How *could* the difference between conscious and unconscious anger come down to something so apparently unimportant as that?’ I take the challenge posed by this question seriously. Indeed, part of what has moved me to pursue a second answer to Q is the aim of coming to better understand how, given this sort of challenge, my first answer could be correct. How is it that gaining, or regaining, an ability to express some emotion or attitude merely by self-ascribing it can make a significant difference in a person’s life? I’m not yet in a position to address this question, but I’ll return to it after I’ve given a second answer to Q, in the final section of this chapter.

### 4. Non-Linguistic Animals

Given the answer to Q that I’ve just been reviewing, we should not expect the distinction between conscious and unconscious psychological states to, as it were, get a grip when we are thinking about non-linguistic animals. After all, it is only when one is considering the psychology of a linguistic creature that it makes sense to distinguish between states that can be expressed in self-ascriptions and states that cannot. And, as a matter of fact, we *don’t* characterize the psychological states of non-linguistic creatures as either conscious or unconscious. Even those of us who are prone to attribute complex thoughts and feelings to our pets do not say things like: ‘My dog gets upset when I pack my suitcase because she has an unconscious fear of being abandoned’, or, ‘Over time, my cat’s distaste for the people who live next door has gone from being unconscious to being conscious’. Nor do we say such things about prelinguistic children. To borrow a phrase (p. 338) that Wittgenstein uses in a related context,<sup>9</sup> when a child learns to attribute mental states to herself, a ‘new joint’ is added to the language-game. We could say that when children learn to self-ascribe their psychological states and events, the *form* of these states and events changes. From then on, they are either conscious, unconscious, or somewhere in between conscious and unconscious.<sup>10</sup> To the extent that some desire of mine is such that I am unable to express it by self-ascribing it, it could be said to suffer from a formal defect or imperfection. But it is no imperfection in Kopi’s desire, e.g. for water that she is unable to express it by self-ascribing it. Kopi’s desires are, unlike mine, neither conscious nor unconscious.<sup>11</sup> This is to say: it makes no sense to describe a desire of hers as either ‘conscious’ or ‘unconscious’. In the shift from brute to fully realized human desire, a new joint is added to the language-game; the psychological is transformed.

## 5. 'What is Gained ... by Substituting Some Words for Some Tears?'

One of the goals that Jonathan Lear sets for himself in his fascinating *A Case for Irony* is to elucidate 'the psychoanalytic meaning of making the unconscious conscious' (Lear 2011: 52). In pursuing this end, he draws on the strand in my work that I've been reviewing and discussing in this chapter. Lear contrasts a case in which someone concludes on the basis of self-observation that he must be angry, even though he isn't 'consciously feeling angry', with a very different case that he describes as follows:

[I]n the midst of a boiling rage, I say "I'm *furious* with you!" In this case, the anger itself is present in the verbal self-ascription of anger that is directed at you. In this case, the self-ascription of anger is itself an *angry* expression. Unlike the former case, I do not have to observe myself to know that I am angry. I just *am angry*; and the form my anger at you takes on this occasion is the angry verbal self-ascription directed at you. In this case, the utterance "I am *furious* with you!" may replace other forms of angry expression.... This is a case, in Finkelstein's terms, in which I am not only conscious of my anger; I am *consciously angry*.

(Lear 2011: 52-3)

Lear goes on to consider the question of how it might be therapeutic, rather than merely disruptive, to find some way of, as it were, taking unconscious anger—anger that has been 'held out of consciousness not simply because it is painful, but because it violates (p. 339) one's sense of who one is' (Lear 2011: 54)—and making it conscious. In this context, he says, 'The question thus becomes whether there could be a process—not *too* disruptive—by which the verbal expression of anger *replaces* (in the Wittgenstein-Finkelstein sense) its unconscious manifestations' (Lear 2011: 54).

*A Case for Irony* comprises a pair of lectures by Lear followed by responses by several other philosophers. In one of these responses, Richard Moran writes:

The idea of "replacing" one mode of expression for another is not perfectly clear to me, particularly in the therapeutic context into which Lear is importing the idea. On a basic level, we might ask: If the two modes of expression are really doing the same work, then what is the *point* of "replacing" one with the other? If we take the notion of replacement literally, then what is gained, for instance, by substituting some words for some tears? Does that mean that the tears are now unnecessary, having been replaced by something else?

(Moran 2011: 113)

It might be tempting to try to answer Moran's question by drawing a distinction between the kind of expressive 'work' that gets done via a self-ascription of sadness and the kind that gets done via crying. But I believe that this would be to draw a line in the wrong place. At least insofar as we're interested in psychoanalysis, there is an important line for

## Making the Unconscious Conscious

---

us to draw between expressions that are, in a very particular way, defective or imperfect and ones that are not. But, as will become apparent, crying can lie on *either* side of this line. Moran's question is useful, I think, insofar as it helps us to see that we need to distinguish cases in which someone's (e.g.) crying constitutes (what may be called) a conscious expression of sadness (or joy or whatever) from ones in which someone unconsciously expresses her (e.g.) sadness via crying. Moran asks what the point is of replacing some tears with some words—that is, of replacing an expression that takes the form of crying with one that takes the form of self-ascription. My own view is that there need not be *any* point in replacing crying with self-ascription. Tears are just fine as long as they constitute *conscious* expression. Psychoanalysis does not aim to replace tears with words. It aims to replace unconscious expressions with conscious ones.

## 6. The Distinction between Conscious and Unconscious Expressions

What does an unconscious expression look like? We already saw an example of one in §1, but I'd like to consider a somewhat different example now:

One afternoon, Max and his wife, Sarah, have a quarrel while assembling hors d'oeuvres for a dinner party that they'll be hosting that evening. Sarah's new boss will be one of their guests, and partly for this reason, she is worried about the party's going well. Although the quarrel subsides after a few minutes, Max continues to dwell on it. Just before guests are scheduled to arrive, he reminds himself that Sarah (p. 340) has been anxious about the party's going smoothly, and he tells himself that it would be petty of him to express the anger that he's still feeling toward her in front of their guests. He resolves to be the perfect co-host, to be cheerful, charming, and useful. In spite of—and perhaps partly because of—this resolution, Max repeatedly expresses his anger during the party, but he remains unaware of doing so. He responds to Sarah's sharp, witty, and critical assessment of a recent movie by suggesting that she simply failed to appreciate what was innovative about it; he makes disparaging jokes about food that Sarah prepared; he tells a story about her brother that she finds embarrassing; and so on. It does not occur to Max that these remarks of his might be connected to that afternoon's quarrel, and if someone were to suggest that there *was* a connection, he would deny it. Nonetheless, it is apparent to all but the drunkest of their friends that Max is annoyed with Sarah. For all that, Max performs the bulk of the serving and most of the straightening up. Indeed, at one point during the evening, he privately congratulates himself on managing to 'put on a good face' for Sarah's sake, in spite of his angry, hurt feelings.

Here, Max unconsciously expresses his anger toward Sarah in passive-aggressive remarks. A noteworthy feature of this particular example of unconscious expression—a feature that is not present in every example—is that here, a person unconsciously expresses a state of mind that is, itself, conscious. (As I've asked you to imagine him, Max

## Making the Unconscious Conscious

---

is consciously angry at Sarah.) This example stands in contrast to one in which someone unconsciously expresses an unconscious state of mind. The story that I told in §1—in which Max expresses his unconscious anger by repeatedly stranding his mother at the airport—constitutes one such example.

Thus far, I've recounted two stories in which Max *unconsciously* expresses anger. Here's one in which he *consciously* expresses pleasure:

Early one evening, Max and Sarah are feeling both hungry and tired, so rather than cooking dinner, they decide to go to a nearby restaurant. Five minutes after they order, Sarah tells a joke, and after a brief pause, Max's face lights up in a smile. Sarah says, 'Are you smiling at my joke?' Max replies: 'No. I thought your joke *was* funny, but I'm smiling because I see the waiter with our food, and I'm really pleased that it's coming out so quickly'.

In this example, Max is able to speak to the question of what his smile expresses in, we could say, a first-personal way. His smile is a conscious expression of his pleasure at the imminent arrival of dinner.

How is conscious expression possible? How is it that Max is able to just say what his smile expresses? How does he know that he's *not* smiling at Sarah's joke? I want to approach these questions by first considering what seems to me an easier sort of case, one in which someone speaks with a kind of first-person authority—not about what psychological state a bit of his behaviour expresses but rather—about what some of his own words mean:

One afternoon, Max and Sarah attend a philosophy department social event. A graduate student says to Max: 'I just met Sarah. Is she a philosopher too?' Max (p. 341) replies: 'She *is* a philosopher, by which I mean: she is moved by, and thinks hard about, philosophical questions. But she's not someone who makes her living by doing philosophy'.

Here, Max first says, 'Sarah *is* a philosopher', and then he provides a gloss on what he means by these words. Imagine that someone were to ask: 'How did Max find out what he meant by the words, "Sarah *is* a philosopher"? Did he first hear himself utter these words and then work out a plausible interpretation of them?' That's the wrong way to think about what's going on in the story. Indeed, the question, 'How did Max find out what he meant when he said, "Sarah *is* a philosopher"?' is itself confused. Max didn't *find out* what the first part of his statement meant and then, in the second part of the statement, report what he had learned. We might want to call the second half of Max's statement an interpretation of, or gloss on, the first half. But it's not the sort of gloss or interpretation that another person might provide. The second half of Max's statement is as much a part of his assessment of Sarah *qua* philosopher as the first is. We should think of the two halves of Max's statement as constituting a kind of unity, a single act. I'll describe them as *co-expressing* his assessment of Sarah *qua* philosopher. If a listener wants to understand this assessment, she needs to consider the two parts of Max's statement (the

## Making the Unconscious Conscious

---

part in which he says, 'Sarah is a philosopher', and the part in which he glosses what he means by these words) as making sense together, as two pieces of a single, extended expression of his thought.

Now consider again the story I told about Max and Sarah at the restaurant. Recall that when Sarah asks Max whether he's smiling at her joke, he replies, 'I'm smiling because I see the waiter with our food, and I'm really pleased that it's coming out so quickly'. Max's smile, I said, is a conscious expression of his pleasure at the imminent arrival of dinner. I want to suggest that we understand Max's gloss on what his smile expresses in the same way that we understood Max's gloss on what he means by the words, 'Sarah is a philosopher'. When Max says what his smile expresses, he is not offering the sort of interpretation of his behaviour that another person might offer. Rather, his smile and his gloss on its significance are two parts of a unified expressive act; together, the smile and the gloss co-express his pleasure at seeing dinner on its way. What I'll call a 'co-expressive gloss' on the psychological significance of, e.g. a smile is not an interpretation of the smile based on observation. It is, rather, a kind of extension of the smile into a linguistic register.

\*\*\*

At this point, I can say what it is that distinguishes conscious expressions from unconscious ones. Someone's expression of a mental state is conscious if she is able to gloss it—i.e. gloss its psychological significance—co-expressively. Someone's expression of a mental state is unconscious if she is unable to gloss it co-expressively. What it means to say that Max's passive-aggressive dinner-party remarks are unconscious expressions of his anger at Sarah is this: they are expressions of anger at Sarah that Max is unable to gloss co-expressively.

(p. 342) It's not impossible that someone should unconsciously express her frustration by slamming a door and, immediately thereafter, accurately interpret the significance of this behaviour in the way that a psychologically astute external observer might. But in such a case, the door-slam and the gloss on it would be separate acts. When someone consciously expresses her frustration by slamming a door, she is able to provide a gloss on the slam, where such a gloss would, itself, be a continuation—a kind of extension or stretching out—of her expressive act.

## 7. Replacement

Earlier in this chapter (§2), I discussed §244 of *Philosophical Investigations*, in which Wittgenstein introduces the idea that when a child learns to talk about his pains, he thereby learns a new way to express them. I noted that §244 concludes in a way that's liable to sound off-key, with Wittgenstein saying to his interlocutor, '[T]he verbal expression of pain replaces crying'. I went on to make a suggestion that I was not yet in a position to spell out, viz. that even if 'replaces' ('ersetzt') seems the wrong word for Wittgenstein to have used in this context, there is a sense in which the kind of crying that

## Making the Unconscious Conscious

---

children do as prelinguistic infants is—when they learn to talk—replaced by another kind of crying. I want to return to that suggestion and try to explain it in light of what I just said about the distinction between conscious and unconscious expressions along with something I said a little while ago, in §4.

In §4, I noted that the distinction between, e.g. anger that can be expressed via self-ascription and anger that cannot be so expressed does not get a grip when one is considering the mind of a non-linguistic or prelinguistic animal. I went on to say that when a child gains the ability to self-ascribe his attitudes and emotions, their form changes. Until then, it doesn't make sense to describe them as conscious or as unconscious. From then on, they are either conscious or unconscious or in between conscious and unconscious. Now, something similar happens when a child gains the ability to co-expressively gloss the psychological significance of his expressions: the form of his *expression* changes. Until then, it doesn't make sense to describe his expressions (of sadness, pleasure, love, desire, anger, etc.) as either conscious or as unconscious. From then on, his expressions are either conscious or unconscious or in between conscious and unconscious. When Kopi expresses a desire to play with me by dropping a Frisbee at my feet, it is no formal imperfection in her expression that she is unable to extend it into a gloss on itself. But if Sarah discovers Max curled up on their bed, sobbing, and if, when she asks him why he is crying, he is completely unable to say, then his crying could be characterized as imperfectly instantiating the form of human expression. Thus, when children learn to talk, one kind of crying does, after all, replace another.

Perhaps this a good place to point out that in this chapter, I am discussing four distinct kinds of replacement: (i) the replacement of prelinguistic desires, fears, etc. with the sort of desires, fears, etc. that are characteristic of linguistic creatures, (ii) the replacement (p. 343) of prelinguistic expressions of desire, fear, etc. with the sort of expressions of desire, fear, etc. that are characteristic of linguistic creatures, (iii) the replacement of unconscious desires, fears, etc. with conscious ones, and (iv) the replacement of unconscious expressions of desire, fear, etc. with conscious ones. A case of type (i) or (ii) does not constitute an example of 'making the unconscious conscious'.<sup>12</sup> Cases of types (iii) and (iv) *are* examples of 'making the unconscious conscious'. A central aim of this chapter is to elucidate the logical relations between cases of types (iii) and (iv)—between, that is, two different ways in which the unconscious may be made conscious.

## 8. Another Answer to Q

In the story that I told about Max and Sarah's dinner party, Max expresses his anger at Sarah via passive-aggressive remarks. As I noted in §6, that story illustrates that it is possible for someone to unconsciously express his conscious anger. Indeed, there is nothing particularly extraordinary about a person's unconsciously expressing either an unconscious emotion or a conscious one. By contrast, we don't find cases that we'd be inclined to describe as ones in which someone *consciously* expresses an *unconscious* state of mind. This is what we should expect, given what I have claimed about the distinction

## Making the Unconscious Conscious

---

between conscious and unconscious expressions. In order for Max to consciously express, e.g. his anger toward his mother, he would need to produce an expression of anger whose psychological significance he could co-expressively gloss. But now: consider a case in which Max expresses his anger toward his mother by not answering the phone when he sees that she's calling. If Max *could* co-expressively gloss this expression—if he could say, 'I'm not answering the phone because I'm angry at my mother', and, thereby, extend his non-verbal expression into a gloss—his anger would be conscious.

Summing up a bit: we've seen that conscious states of mind may be expressed either consciously or unconsciously. But unconscious states of mind can be expressed only unconsciously. There is, as it were, no grammatical or logical space for an unconscious emotion or attitude to be consciously expressed. A little reflection on this point yields the new answer to Q that I promised at the start of this chapter. In §3, I quoted the following statement from *Expression and the Inner*:

*Someone's mental state is conscious if he has an ability to express it merely by self-ascribing it. If he lacks such an ability with respect to one of his mental states, it is unconscious.*

(Finkelstein 2003: 120)

(p. 344) Let's call that first answer to Q 'A1'. What follows is a second answer; call it 'A2':

*Someone's mental state is conscious if she has an ability to consciously express it. If she lacks this ability with respect to one of her mental states, it is unconscious.*

A2 might sound as if it involves an uncomfortably tight circle, so I should point out that although I am here, as it were, defining a sort of 'consciousness' in terms of a sort of 'consciousness', the latter sort of consciousness has been further defined in terms of co-expressive glossing.<sup>13</sup>

Although A1 and A2 may be thought of as providing two different criteria for a mental state's counting as conscious, if they both are true—as I believe they are—then any mental state that meets either one of these criteria will meet the other one as well. This is to say: (i) if someone is able to express his, e.g. anger merely by self-ascribing it, then he will be able to consciously express it. And (ii) if someone is able to consciously express, e.g. his anger, then he will be able to express it merely by self-ascribing it.<sup>14</sup> In what remains of this chapter, I'll be particularly interested in a point that is implied by (ii), viz. that if someone is *unable* to express his, e.g. anger merely by self-ascribing it, then he will be *unable* to consciously express it.

\*\*\*

Given the answer to Q that I summarized in §3 and that is encapsulated in A1, making an unconscious emotion or attitude conscious is a matter of acquiring or regaining an ability to express it merely by self-ascribing it. It is such an ability that, I said, Max still (p. 345) lacks even after he's concluded on the basis of behavioural evidence that he is angry at

## Making the Unconscious Conscious

---

his mother. At the end of §3, I introduced, and then left hanging, a question concerning how this could be the right thing to say about what it takes for an unconscious state of mind to become conscious. Someone might put this question to me as follows: 'As you have characterized him, Max is able to express his unconscious anger at his mother by stranding her at the airport, by misplacing his phone when he knows that she is likely to call, and, presumably, by doing a wide variety of other things that are liable to annoy her. Given the long list of ways in which he can *already* express this anger, why should his becoming able to express it in *one more way*—by self-ascribing it—be significant? How could the difference between unconscious and conscious anger come down to *that*?'

I am now in a better position than I was in §3 to address the challenge presented by this question. A couple of paragraphs back, I noted that putting A1 and A2 together yields a point that I put as follows: 'if someone is *unable* to express his, e.g. anger merely by self-ascribing it, then he will be *unable* to consciously express it'. I suggest that this point ought to help us understand why it is that the ability to express one's own state of mind in a self-ascription of it should *not* be considered just one expressive ability on a long list of such abilities. Until Max is able to express his anger at his mother merely by self-ascribing it, there is no way for him to consciously express it at all—either by telling his mother that he is angry at her or by refusing to let her stay with him and Sarah when she visits or by choosing not to answer the phone when he sees that she is calling.<sup>15</sup> If Max were to gain the ability to express his anger at his mother merely by self-ascribing it, this would, as it were, open up the possibility of his consciously expressing it in these and myriad other ways. If only for this reason, such a shift would be significant; it would matter to Max's life.

For all that, acquiring an ability to express his anger toward his mother in a self-ascription of it would not guarantee that Max's subsequent expressions of this anger would be conscious rather than unconscious. As we've seen, a conscious state of mind may be expressed *either* consciously or unconsciously. At the end of §5, I said that psychoanalysis aims—not to replace tears with words but—to replace unconscious expressions with conscious ones. It is perhaps disappointing that acquiring (or regaining) an ability to express an emotion or attitude in a self-ascription of it does not ensure that it won't continue to be expressed unconsciously. Still, the acquisition of such an ability is a necessary step toward replacing unconscious expressions with conscious ones.<sup>16</sup>

## References

Finkelstein, D. H. (1999). 'On the distinction between conscious and unconscious states of mind'. *American Philosophical Quarterly* 36(2) (April): 79–100.

Finkelstein, D. H. (2003). *Expression and the Inner*. Cambridge, MA: Harvard University Press.

Finkelstein, D. H. (2010). 'Expression and avowal'. In K. Jolley (ed.), *Wittgenstein: Key Concepts* (pp. 185–198). Durham: Acumen Press.

## Making the Unconscious Conscious

---

Kluger, J. (2017). *The Animal Mind*. New York: Time Inc. Books.

Lear, J. (2011). *A Case for Irony*. Cambridge, MA: Harvard University Press.

Moran, R. (2011). 'Psychoanalysis and the limits of reflection'. In J. Lear, *A Case for Irony* (pp. 103–114). Cambridge, MA: Harvard University Press.

Wittgenstein, L. (1953). *Philosophical Investigations*, trans. G. E. M. Anscombe. Oxford: Basil Blackwell.

Wittgenstein, L. (1967). *Zettel*, trans. G. E. M. Anscombe, eds G. E. M. Anscombe and G. H. von Wright. Oxford: Basil Blackwell.

Wittgenstein, L. (1992). *Last Writings on the Philosophy of Psychology Volume 2*, trans. C. G. Luckhardt and M. A. E. Aue, eds G. H. von Wright and H. Nyman. Oxford: Blackwell.

### Notes:

(<sup>1</sup>) See, e.g. Finkelstein 1999 and Finkelstein 2003.

(<sup>2</sup>) Although Lear is interested in how my answer to Q might bear on our understanding of psychoanalysis, it's worth mentioning that neither Q nor the answer to it that I've given is *about* psychoanalysis as such. (Freud was not the first writer to recognize that we are sometimes moved by unconscious emotions, preferences, and desires. I take it that the title character in Jane Austen's *Emma*, who realizes at the end of the novel that she's been in love with Mr. Knightly for quite a while, presents at least as good an example of someone whose behaviour is motivated by unconscious states of mind as does, say, Freud's Rat Man.)

(<sup>3</sup>) I believe that many theorists of 'consciousness' fall into confusion by failing to distinguish between 'conscious of x' and 'x's being conscious'. For a discussion of this point, see Finkelstein 2003, ch. 1.

(<sup>4</sup>) In her translation of *Philosophical Investigations*, G. E. M. Anscombe sometimes renders 'äußerung' as 'expression' and sometimes as 'manifestation'. Here, I've changed her 'manifestation' to 'expression'.

(<sup>5</sup>) It would take me beyond the scope of what I aim to do in this chapter to explain why Wittgenstein thinks this is misguided. Stepping back from Wittgenstein, I would suggest that any such position will suffer from the conflation that I mentioned in n. 3.

(<sup>6</sup>) One difference is that wincing, moaning, and crying (regardless of whether or not they are produced by a human being who can already talk) are not truth-apt, while self-ascriptions are. Wittgenstein has often been read as wanting to deny this difference and claim that psychological self-ascriptions are neither true nor false. This reading is, I think, both uncharitable and unsupported by the relevant texts. (For a discussion of this point, see Finkelstein 2010.)

## Making the Unconscious Conscious

---

(<sup>7</sup>) ‘What if someone has a phobia that prevents him from ever speaking about being angry? Couldn’t he nevertheless be consciously angry?’ As I understand the concept of expression, not every expression must be a *public* manifestation. (It seems to me that on more than one occasion, I have: (i) tried to tiptoe past a sleeping person, (ii) stubbed my toe, and, finally, (iii) expressed the resultant pain in my toe by *thinking*, ‘Ow! That really, really hurts!’) It’s consistent with my understanding of expression to allow that we *sometimes* express our states of mind by self-ascribing them in thought, rather than out loud. So yes; someone who has a phobia about *saying* that he’s angry can, on my view, nevertheless be consciously angry.

(<sup>8</sup>) Imagine someone’s replying to what I said in this sentence as follows: ‘If a person expresses his, e.g. anger by self-ascribing it, then what his self-ascription “gives voice to” is just this anger and *not*, as you suggest here, a judgement about it’. In order to bracket this concern, my former self might have rewritten this sentence so that it ended as follows: ‘but simply the fact that he is sincerely addressing a question about his own state of mind’.

(<sup>9</sup>) Wittgenstein 1967: §§422–6.

(<sup>10</sup>) I argue that we should think of conscious and unconscious states of mind as lying at two ends of a continuum in Finkelstein 2003: §5.5.

(<sup>11</sup>) This isn’t to deny that Kopi might be very conscious *of*, e.g. a piece of cheese that’s sitting on a plate on a table just above her head. Nor is it to deny that she might bang her head—perhaps in pursuit of that cheese—and momentarily *lose consciousness*. (In a special edition of *Time* called *The Animal Mind*, Steven Pinker is quoted as saying, ‘It would be perverse to deny consciousness to mammals’ (Kluger 2017: 11). When I say that Kopi’s states of mind are neither conscious nor unconscious, I don’t take myself to be disagreeing with Pinker’s remark.)

(<sup>12</sup>) I’ve characterized cases of types (i) and (ii) as involving changes in form—*from* states of mind and expressions that cannot be described as either conscious or unconscious *to* states of mind and expressions that can be so described. Thus, cases of these types are examples—not of ‘making the unconscious conscious’ but rather—of ‘making what is neither conscious nor unconscious either conscious or unconscious’.

(<sup>13</sup>) I am not claiming that there won’t turn out to be *some* sort of circularity involved in one or both of my characterizations of the distinction between conscious and unconscious states of mind. My aim is not to provide a reductive account of consciousness. I don’t mind circles, as long as they aren’t so tight as to be uninteresting.

(<sup>14</sup>) If it still seems unclear that (i) and (ii) both obtain, what follows might make it seem less so. Let’s begin with (ii). Imagine that Max slams his office door and that this is (for him, in this situation) a conscious expression of anger that he’s feeling toward a colleague named Jeff. What it means for this to be a *conscious* expression of anger is that he can co-expressively gloss its psychological significance; he can respond to a question like, ‘Why

## Making the Unconscious Conscious

---

did you slam the door?’ with an answer like, ‘I slammed it because I’m angry at Jeff’—and he can do this in such a way that the gloss is a kind of extension of (rather than an observation report about) the initial expressive act (the door-slam). Now, if Max can do that, then he is (*obviously*, I want to say) able to express his anger merely by self-ascribing it, i.e. merely by saying, ‘I’m angry at Jeff’. So, if Max’s anger counts as conscious according to A2, it will also count as conscious according to A1. Thus, (ii) would seem to be true.

Now let’s consider (i). Imagine that Max is able to express his anger at Jeff merely by saying, ‘I’m angry at Jeff’. Such a self-ascription would *itself* be a conscious expression. This is to say: Max would have the ability to co-expressively gloss *its* psychological significance. Suppose that Max expresses his anger at Jeff by saying to Sarah, ‘I’m angry at Jeff’. Sarah thinks (for reasons with which we needn’t concern ourselves) that in so saying, Max might be trying to make a joke. She replies, ‘Are you joking? I’m not getting what you mean’. Here, Max would be in a position to reply with a co-expressive gloss on his initial expression, as follows: ‘When I said, “I’m angry at Jeff”, I wasn’t making a joke; I was actually expressing anger that I’m feeling’. So, if Max’s anger meets the criterion for consciousness in A1, it will also meet the criterion in A2. Thus, (i) would also seem to be true.

(<sup>15</sup>) While his anger toward his mother remains unconscious, Max might unconsciously express it by not answering the phone when he sees that she is calling. But until he can express his anger merely by self-ascribing it, there is no possibility of his *consciously* expressing it by not answering the phone when he sees that she is calling.

(<sup>16</sup>) Versions of this chapter were presented at the University of Murcia (Workshop on Contemporary Rationalist and Expressivist Approaches to Self-knowledge, October 2016) and at the University of Chicago (2016 Fall Colloquium, Department of Philosophy, November 2016). I’m grateful to members of both audiences for their questions and comments, especially Matthew Boyle, Jonathan Lear, Ángel García Rodríguez, Ryan Simonelli, and Anubav Vasudevan. In addition, I’m indebted to Agnes Callard, Matthias Haase, and Malte Willer for reading and commenting on drafts.

**David H. Finkelstein**

Department of Philosophy, University of Chicago, USA