



## Chapter 4

# Knowing What It's Like

Joseph Levine

Externalist Representationalism about Qualia (ERQ) is a very tempting position.<sup>1</sup> Those seemingly ineffable sounds, sights, tastes and smells that populate our stream of conscious experience are maddeningly difficult to incorporate into the realm of natural, physical phenomena. On the one hand, it's really hard to see how the reddish-orange character of seeing a sunset could be reducible to, or completely explicable in terms of, the neurological processes of the visual system. How does neural firing, no matter how complex, amount to a reddish-orange experience? On the other hand, analysing the qualitative character of experience in purely functional terms presents its own difficulties. Again, how does a formal property such as occupying a particular causal role or a point in a multidimensional space, amount to the seemingly quite concrete reddish-orange character of seeing a sunset? True, both traditional functionalism and psychoneural reductionism have their adherents,<sup>2</sup> but their inadequacies are clearly felt by many philosophers.

Enter ERQ. ERQ vindicates the feeling that there really are these qualities encountered in experience – the reddish-orange character of the sunset, the bitter taste of a lemon. But there's nothing threatening to physicalism in acknowledging their existence because they are just the properties of the distal objects of perception that physical accounts of nature are intended to explain. It's the actual colour of the sky or the chemical composition of the lemon that are the qualities encountered. It thus looks like we can eat our cake and have it too.

Of course, qualia can't be straightforwardly identified with the properties of external objects, with colours and chemical properties of fruit. For one thing, we have to contend with hallucinations. We can 'see' sunsets that aren't there and 'taste' imaginary lemons. Also, experiences are mental events, and it's hard to see how the actual colour of the sunset or the chemical composition of the lemon can 'get inside' the head. But this isn't really a problem, since the qualitative character of the experience itself, according to ERQ, is not literally the colour or taste we're aware of, but rather the property of representing that colour or taste. Just as the representationalist solution to the problem of how unactualized goals can motivate actions is to appeal to actual, inside-the-head representations of goals as the efficacious agents, so too ERQ brings the sensible qualities of external objects inside the head by way of being represented. All in all, a very elegant solution.

Except that it doesn't work. The basic problem is that how the outside world looks, tastes and smells is too much a function of what's going on inside us. One way to see this is to consider various qualia inversion scenarios, such as the famous 'inverted spectrum hypothesis'.<sup>3</sup> If indeed it's possible that the character of my

experience when normally viewing a ripe tomato is phenomenologically like the character of your experience when viewing a ripe cucumber, then, on certain plausible assumptions about what determines the external representational contents of visual perceptions, our qualia can't be reduced to these contents. But, of course, whether or not qualia inversions really constitute coherent possibilities is a matter of much controversy.<sup>4</sup> So I will focus instead on another argument. I will argue that there is a conflict between ERQ and a plausible condition on self-knowledge of qualitative experience.

Initially, the problem is not hard to motivate. As I look at the reddish-orange light of the sunset, I know what it's like for me. There's a way I can be secure in that knowledge despite harbouring doubts about whether the sun is really setting, whether I'm dreaming, or even whether I'm a brain in a vat. One needn't be committed to a full-blown doctrine of incorrigibility in order to acknowledge a difference between the security of our knowledge of what our experience is like and the security of our knowledge of what's going on outside us. Yet, it would seem, if what our experience is like is metaphysically determined by what's going on outside us, this epistemic difference wouldn't exist.

The problem isn't that ERQ cannot accommodate the intuition that we can be more certain of how things seems to us, or look to us, than of how they really are. Remember, ERQ has no trouble handling hallucinations. The conflict between ERQ and self-knowledge only even appears to arise once we append to ERQ a theory of perceptual representation: an account of what determines the representational content of a percept.

It seems to me that there have only been two theories of representation for percepts ever proposed (at least in modern philosophy) and I can't think of any alternative to them: percepts represent the qualities of objects either by resembling them, or by standing in an appropriate causal relation to them.<sup>5</sup> What's more, resemblance seems to be a non-starter, at least for physicalists (on whose behalf, after all, ERQ was proposed to begin with). However, once you combine some sort of causal theory of representation with ERQ, then you seem to get the conflict with self-knowledge. For can't you know about what it's like for you to see, hear, taste and smell without that knowledge being somehow held hostage to facts about the properties of the external objects with which you are (or have been, or your ancestors have been) causally interacting?

ERQ doesn't go down that easily though. In order to see what the real problem is, it's important to compare and contrast ERQ with an externalist theory of content for cognitive states such as thoughts, beliefs, and desires – call it 'standard content externalism' (SCE). For there has been a longstanding discussion about the question whether SCE is compatible with a plausible theory of what we can know about the contents of our own thoughts. Given that many philosophers believe that there is no ultimate incompatibility there, it certainly seems promising to apply their arguments to the alleged incompatibility between ERQ and self-knowledge.

Suppose I'm thinking that I'd like a drink of water. We can characterize my desire as having the content [I have a drink of water]. Assuming there is nothing funny going on, my having this desire, with this content, is something that I know. If you ask me what I'm after, I'll tell you 'a drink of water'. While there clearly can be desires of which I'm not aware, this isn't one of them.

However, on SCE, the content of my desire for a drink of water is in fact a desire for a drink of  $H_2O$ . Furthermore, my twin on Twin Earth, who is individualistically identical to me, actually has a desire with a different content – he desires a drink of XYZ. I cannot, from the inside, distinguish having a desire with the one content from having a desire with the other. Hence, in a certain sense I don't know what it is I desire. But this is absurd. Of course I know what I want – a drink of water.

One option is to type-identify thought contents narrowly, so that I and my twin count as having desires of the same content type. If there is no difference in content between my twin's desire and my own, then there's no cause to indict me of inappropriate self-ignorance. However, this move relies on a robust notion of narrow content, and many philosophers have trouble with this.<sup>6</sup>

Fortunately for externalists, there is another option. Essentially, the idea is to let the contents of my higher-order, self-regarding cognitive states be determined by the contents of the lower-order ones on which they are directed.<sup>7</sup> So, yes, the content of my desire is [I have a drink of  $H_2O$ ] if I'm on Earth and [I have a drink of XYZ] if I'm on Twin Earth. Also, it's true that I can't tell, merely from what's going on inside, which one of the two planets I live on. But that doesn't entail that I'm ignorant of the content of my desire. Why? Because what I know is that I desire a drink of water, and whatever 'water' means in the first-order expression of my desire it also means in the expression of any higher-order state directed on my desire. It's certainly not part of the externalist position that in order to know something about water I have to know that it's identical to  $H_2O$  (or, more precisely, know that ' $H_2O$ ' applies to it). Similarly, in order to know what the content of my desire is I don't have to know every possible description that expresses that content. One such description is sufficient.

Let's return now to ERQ. According to ERQ, the reddish-orange character of my visual experience of a sunset is reducible to the content (with respect to colour) of that experience. That is, the qualitative character is the property of representing RO, where 'RO' picks out whatever property in the world is the sky's having a certain reddish-orange glow. We can suppose that it is some quite complicated physical property of the light emanating from that part of the sky. Thus the qualitative character of my experience just is its representing that quite complicated physical property, RO.

The problem for ERQ was supposed to be this. If my having an experience with a reddish-orange character just is having one that represents RO, then, if I don't know anything about the physical property RO, it seems I won't know anything about the character of my experience. But, of course, I do know what my experience is like even if I know nothing of the physics of light or the psychophysics of vision. So therefore ERQ is in trouble.

However, we can see that the externalist strategy above can apply straightforwardly here as well. What I know about my experience is that it presents me with a reddish-orange field. Whatever means by which this self-regarding knowledge is represented is parasitic on the representation by which the experience itself is expressed (or tokened). Of course, I do not know the complicated physical description of the light emanating from the horizon, but I do know of that very physical property under the description 'reddish-orange' (or by whatever means of representation is used by the visual system). I employ that same representation in characterizing my experience to myself. Thus, it seems, ERQ entails no inappropriate self-ignorance on my part.

experience when normally viewing a ripe tomato is phenomenologically like the character of your experience when viewing a ripe cucumber, then, on certain plausible assumptions about what determines the external representational contents of visual perceptions, our qualia can't be reduced to these contents. But, of course, whether or not qualia inversions really constitute coherent possibilities is a matter of much controversy.<sup>4</sup> So I will focus instead on another argument. I will argue that there is a conflict between ERQ and a plausible condition on self-knowledge of qualitative experience.

Initially, the problem is not hard to motivate. As I look at the reddish-orange light of the sunset, I know what it's like for me. There's a way I can be secure in that knowledge despite harbouring doubts about whether the sun is really setting, whether I'm dreaming, or even whether I'm a brain in a vat. One needn't be committed to a full-blown doctrine of incorrigibility in order to acknowledge a difference between the security of our knowledge of what our experience is like and the security of our knowledge of what's going on outside us. Yet, it would seem, if what our experience is like is metaphysically determined by what's going on outside us, this epistemic difference wouldn't exist.

The problem isn't that ERQ cannot accommodate the intuition that we can be more certain of how things seems to us, or look to us, than of how they really are. Remember, ERQ has no trouble handling hallucinations. The conflict between ERQ and self-knowledge only even appears to arise once we append to ERQ a theory of perceptual representation: an account of what determines the representational content of a percept.

It seems to me that there have only been two theories of representation for percepts ever proposed (at least in modern philosophy) and I can't think of any alternative to them: percepts represent the qualities of objects either by resembling them, or by standing in an appropriate causal relation to them.<sup>5</sup> What's more, resemblance seems to be a non-starter, at least for physicalists (on whose behalf, after all, ERQ was proposed to begin with). However, once you combine some sort of causal theory of representation with ERQ, then you seem to get the conflict with self-knowledge. For can't you know about what it's like for you to see, hear, taste and smell without that knowledge being somehow held hostage to facts about the properties of the external objects with which you are (or have been, or your ancestors have been) causally interacting?

ERQ doesn't go down that easily though. In order to see what the real problem is, it's important to compare and contrast ERQ with an externalist theory of content for cognitive states such as thoughts, beliefs, and desires – call it 'standard content externalism' (SCE). For there has been a longstanding discussion about the question whether SCE is compatible with a plausible theory of what we can know about the contents of our own thoughts. Given that many philosophers believe that there is no ultimate incompatibility there, it certainly seems promising to apply their arguments to the alleged incompatibility between ERQ and self-knowledge.

Suppose I'm thinking that I'd like a drink of water. We can characterize my desire as having the content [I have a drink of water]. Assuming there is nothing funny going on, my having this desire, with this content, is something that I know. If you ask me what I'm after, I'll tell you 'a drink of water'. While there clearly can be desires of which I'm not aware, this isn't one of them.

However, on SCE, the content of my desire for a drink of water is in fact a desire for a drink of  $H_2O$ . Furthermore, my twin on Twin Earth, who is individualistically identical to me, actually has a desire with a different content – he desires a drink of XYZ. I cannot, from the inside, distinguish having a desire with the one content from having a desire with the other. Hence, in a certain sense I don't know what it is I desire. But this is absurd. Of course I know what I want – a drink of water.

One option is to type-identify thought contents narrowly, so that I and my twin count as having desires of the same content type. If there is no difference in content between my twin's desire and my own, then there's no cause to indict me of inappropriate self-ignorance. However, this move relies on a robust notion of narrow content, and many philosophers have trouble with this.<sup>6</sup>

Fortunately for externalists, there is another option. Essentially, the idea is to let the contents of my higher-order, self-regarding cognitive states be determined by the contents of the lower-order ones on which they are directed.<sup>7</sup> So, yes, the content of my desire is [I have a drink of  $H_2O$ ] if I'm on Earth and [I have a drink of XYZ] if I'm on Twin Earth. Also, it's true that I can't tell, merely from what's going on inside, which one of the two planets I live on. But that doesn't entail that I'm ignorant of the content of my desire. Why? Because what I know is that I desire a drink of water, and whatever 'water' means in the first-order expression of my desire it also means in the expression of any higher-order state directed on my desire. It's certainly not part of the externalist position that in order to know something about water I have to know that it's identical to  $H_2O$  (or, more precisely, know that ' $H_2O$ ' applies to it). Similarly, in order to know what the content of my desire is I don't have to know every possible description that expresses that content. One such description is sufficient.

Let's return now to ERQ. According to ERQ, the reddish-orange character of my visual experience of a sunset is reducible to the content (with respect to colour) of that experience. That is, the qualitative character is the property of representing RO, where 'RO' picks out whatever property in the world is the sky's having a certain reddish-orange glow. We can suppose that it is some quite complicated physical property of the light emanating from that part of the sky. Thus the qualitative character of my experience just is its representing that quite complicated physical property, RO.

The problem for ERQ was supposed to be this. If my having an experience with a reddish-orange character just is having one that represents RO, then, if I don't know anything about the physical property RO, it seems I won't know anything about the character of my experience. But, of course, I do know what my experience is like even if I know nothing of the physics of light or the psychophysics of vision. So therefore ERQ is in trouble.

However, we can see that the externalist strategy above can apply straightforwardly here as well. What I know about my experience is that it presents me with a reddish-orange field. Whatever means by which this self-regarding knowledge is represented is parasitic on the representation by which the experience itself is expressed (or tokened). Of course, I do not know the complicated physical description of the light emanating from the horizon, but I do know of that very physical property under the description 'reddish-orange' (or by whatever means of representation is used by the visual system). I employ that same representation in characterizing my experience to myself. Thus, it seems, ERQ entails no inappropriate self-ignorance on my part.

However, despite the apparent escape, I think, in the end, there really is a conflict between ERQ and self-knowledge, a kind of conflict that does not arise between SCE and self-knowledge. Before describing the conflict, let me first say something about a crucial difference between SCE and ERQ – one that explains why the conflict to be described below between ERQ and self-knowledge does not afflict SCE. The difference between SCE and ERQ is not a matter of their different domains of application – thought in the former case and phenomenal experience in the latter. Rather, it is a difference in the use to which externalism is put in the two theories. SCE is a theory about the *contents* of thoughts (and other propositional attitudes). That is, it type-individuates the *contents* of thoughts by reference to external factors. Now this is quite significant if you think that psychological generalizations apply to beliefs and desires by virtue of their contents.<sup>8</sup> Still, SCE is a theory of the individuation of contents of thoughts, *not* a theory of the individuation of thoughts themselves.

ERQ, on the other hand, is a theory that combines externalism about content with a reduction of phenomenal character to content. This second element of the theory, the reduction of phenomenal character to content, when combined with externalism about content, is what causes the conflict with self-knowledge. Thus ERQ stands in contrast to SCE, which, by not reducing thought to its content, can avoid a conflict with self-knowledge.

So what's the nature of the conflict? It has to do with the existence (or even possibility) of Frege cases. As is well known, Frege cases are instances where, because we apprehend the same object (or property) in different ways, we are ignorant of the relevant identity. I think that Hesperus and Phosphorus are different heavenly bodies, although it turns out that they are one and the same planet, Venus. Oedipus thinks he's marrying Jocasta but not marrying his mother – oops, he's wrong about the latter!

What's crucial about Frege cases, and what exercised Frege about them, is that the ignorance they involve is purely empirical; it involves no internal failure of rationality, or, to put it more psychologically than Frege would have liked, no failure of cognitive mechanisms. No matter how well one's internal mechanisms are functioning, such judgements of non-identity are hostage to future discoveries concerning extra-mental matters of fact. Oedipus might have been as justified and certain as he could be that marrying Jocasta didn't amount to marrying his mother, but, alas, he was wrong all the same.

On the other hand, it does seem that, on the assumption nothing bizarre is going on inside, Oedipus can be certain about one thing: he doesn't think Jocasta is his mother. True, we can construct a Freudian story on which deep down he actually knows the truth about his future bride but is repressing it. I don't deny this possibility. My point is only this. Whether or not the thought that Jocasta is the one he wants to marry is the same as the thought that his mother is the one he wants to marry, and that he thinks the one and not the other, is not a judgment that is hostage to the facts about Jocasta herself; his judgement that these are different thoughts isn't rendered false merely by virtue of the fact that Jocasta is indeed his mother. He can know, based merely on what's going on inside his mind, that two thoughts are distinct, even though he can't know on that basis that their truth-conditions are.

Similarly, I may be unable to tell, merely from within, whether my drinking water and my drinking H<sub>2</sub>O are one and the same state, but I can certainly tell that thinking I'm drinking water isn't the same as thinking I'm drinking H<sub>2</sub>O. Externalism may

entail that the identities of my thoughts are partially determined by extra-mental facts, so that in some sense I can't know merely from within the precise identities of my thoughts.<sup>9</sup> But I don't have to know all the facts that go into determining the identities of two thoughts in order to know they are distinct. Their distinctness, I claim, is a fact I can establish on grounds wholly from within.

Now some may think that it follows from SCE that this isn't so, but dispelling this impression was the point of my remarks above about the limited nature of SCE. According to SCE, my two thoughts – that I'm drinking water and that I'm drinking H<sub>2</sub>O – share the same content. But just because they share the same content doesn't mean they are tokens of the same thought type – there's more to a thought than its content, namely the vehicle which expresses that content. So I can tell that my thought that I'm drinking water is not of the same type as the thought that I'm drinking H<sub>2</sub>O, not because I can tell that they have a different content, but because I can distinguish, merely from what's going on inside, that they involve different vehicles. Nothing I may find out subsequently about water and its relation to H<sub>2</sub>O is going to upset the judgement that these are different thoughts.<sup>10</sup>

What emerges from our discussion of Frege cases are two principles: first, with respect to the mind–world relation, there can be no internal guarantee against Frege cases. It can seem 'clear and distinct' to me that *a* is not identical to *b*, and yet, through no epistemic fault of my own, I'm just wrong. Second, my judgements to the effect that two mental states are of distinct types are not subject to this sort of correction from outside. I can know in a definitive way, merely from what's going on inside my mind, not that *a* and *b* differ, but at least that the relevant thoughts involved are different. With these points firmly in mind, let's turn back now to consideration of qualia.

Suppose I'm having experiences of what I judge to be two distinct qualitative types; say one is reddish and the other is greenish. In line with the two principles just presented, I maintain two claims. First, it is absurd to think that my judgement that these two experiences are of distinct type could be shown to be mistaken – that these really aren't different experiences – by appeal to facts external to my mind. I'm not arguing that there couldn't be some bizarre internal cognitive glitch that causes a mistaken judgement about the identity and difference of experiential types. But supposing that everything is functioning internally as it should, and I'm having two experiences simultaneously that I judge clearly to be different, I contend that it's absurd to think this judgement is hostage to facts outside me.

Second, the epistemic situation is drastically different with respect to the external representational contents of experiences. No matter how well my internal cognitive processes are functioning, my judgement that two representations differ in content – that this is not the same thing (or property) as that – is always subject to correction from outside. Thus, for all I know, my thought, or impression, that the distal properties represented by these two experiences are distinct could be wrong. It could be that the surfaces that appear to me greenish and the surfaces that appear to me reddish are really the same.<sup>11</sup> However, if qualitative character is reducible to external representational content, it follows that my judgement that these two experiences are different in character really is subject to external correction. Hence we have a conflict between ERQ and self-knowledge of qualia.

Dretske (1995) challenges the internalist about qualia, who accepts SCE to justify rejecting externalism about qualia while accepting externalism about thought contents.



We've seen above that there is a clear answer to this challenge. SCE does not require that one pin any difference in the identities of thoughts on a difference in their contents; a difference in vehicles also entails a difference in thoughts. Thus if one judges that one is entertaining two distinct thoughts, and nothing bizarre is going on inside, then one is entertaining distinct thoughts. Finding out that they have the same content after all isn't going to require revision of the original judgement.

ERQ, on the other hand, reduces qualitative character to external representational content. So if we do find out that what appeared to us quite clearly to be experiences of distinct qualitative types turn out to have the same external representational content, then we must admit we were wrong. Advocates of ERQ needn't deny, of course, that experiences have vehicles. After all, something has to bear the representational properties. But what they do deny is that qualitative character – what one's experience is like – is determined (even in part) by the identity of the vehicle. Well, if qualitative character is not determined by the vehicle (at least partly), then a difference in vehicle cannot account for a difference in qualitative character. So that's why ERQ creates a conflict with self-knowledge that SCE does not.

I want to emphasize again at this point the difference between my argument here and the sort of conflict with self-knowledge discussed earlier. If one is worried that externalism entails one doesn't know what one is thinking or experiencing, then indeed the same sort of reply to this worry works for both SCE and ERQ. It's sufficient to point out that whatever ambiguity attaches to one's original thought, or experience, regarding its external reference, also attaches to any mental state expressing one's knowledge of the content of one's thought or experience. But my argument does not rely on one's ability to determine the precise identity of one's thought or experience. Rather, it relies on one's ability to detect sameness and difference, whatever the ultimate verdict is on precise identity. This ability we clearly have 'from within', and it would be compromised if the type in question, whether it be thought or experience, were completely determined externalistically.

In the case of SCE, as we've seen, there is room to account for the ability to definitively determine difference by appeal to the vehicles of thought. So too, therefore, should an account of qualitative character allow that a difference in vehicle (whatever that may turn out to be – more on this in a bit) constitutes a difference in qualitative character. Since we can detect a difference in qualitative character from within, and the judgement is not hostage to external facts, then what determines the difference must be grounded on what's within. SCE respects this principle (with respect to thought); ERQ does not.

Now, one might object here that ERQ could be supplemented so as to avoid the conflict with self-knowledge. That is, one could adopt the position that qualitative character was a function of both content and vehicle, so that a difference in one factor is sufficient to make a difference in qualitative character. This would clearly solve the problem, but it would undermine ERQ. The whole idea was to show how what seemed so difficult about the qualitative character of experience could be accounted for within a physicalist framework by reducing qualia to the external qualities of sensible objects. But if it's possible for two states to differ qualitatively even though they do not differ in content, then we've lost the principal advantage of ERQ. For now we have to explain how a difference in vehicle can make a difference in qualitative character.

As mentioned at the start of the essay, there seem to be basically two physicalist options for what determines vehicle identity: either physiological type or functional role. The problems with both these options have already been mentioned, and they provide a major part of the motivation for ERQ. I won't here discuss the question whether a theory of qualia as vehicles of sensory representation can be made to work, except to deal with one issue. While it may be obvious that one's internal monitors could judge physiological identity straightforwardly, one might think that functional identity is another matter. After all, a functional state is defined by its dispositions to interact with lots of other states, so it could seem no easier to determine, in a manner not hostage to further facts, that one's current state satisfies a certain functional definition, than it is to determine its external content. So, at least if one construes the notion of a representational vehicle functionally, it's not obvious that the appeal to vehicles helps.

But remember that it's not vehicle identity *per se* that matters, but vehicle difference. On this question, in fact, we have a kind of self-certifying situation. If I judge (again, with no malfunctions going on) that two states are distinct, *ipso facto* they play distinct functional roles (whatever those roles happen to be in the end). That is, it's definitive of a difference in functional role between vehicles *a* and *b* that I accept  $a = a$  but not  $a = b$ . My very judging them to be different makes them different. So if we were to adopt a functional role theory of qualitative character, the problem from Frege cases would not trouble us.

So far, I've argued that ERQ entails that we couldn't know, merely from inside, that two simultaneous experiences were of distinct qualitative type, and yet clearly we can know this. However, one might very well deny that the sort of situation on which my argument relies could in fact happen. After all, I need a case where two (what appear to be) qualitatively distinct experiences share the same external content. Why think that this can happen? In particular, if one adopts a causal/covariational account of the content of sensory states, how could it happen that the very same distal property gives rise to distinct sensory experiences?

In the case of thoughts, one might argue, there are intermediaries that can account for the standard Frege cases. One sees Hesperus in the evening and Phosphorus in the morning. One identifies water by looking at it or drinking it; one identifies H<sub>2</sub>O by fancy lab tests. Whether or not these differences constitute a difference in sense, or narrow content, as the Fregean would have it, they still account for how two distinct vehicles could hook up to the same referent. But given the immediacy of the relation between the sensible qualities of objects and the sensory detectors in us, it's not clear that there's room for Frege cases to arise.

My response to this objection is twofold. First, in other essays<sup>12</sup> I've tried to construct cases that I think fit the bill. For instance, imagine a creature with eyes on the sides of its head, like a fish. The creature has a head that doesn't turn and lives on a track so it can only move back and forth in a line. The idea is that it can see half its world with one eye, and the other half with the other. Now add to the scenario that the visual systems of the two eyes are inverted with respect to each other. Exposure to red things through one eye causes experiences like those that are caused by exposure to green things through the other eye.

The creature, while looking at red things through both eyes, will judge that it is having experiences of different types, and that in fact the two colours it's seeing are

different. It might one day come to realize, through empirical investigation (or testimony from Earthling visitors) that what looks these two different ways through its two eyes is really the same colour. I don't see any grounds for claiming that one of the eyes 'gets it right'; they both represent the same colour, at least on a causal/covariational account. Yet, on discovering that the colour is the same, the creatures aren't going to now deny that their experiences are different.

I find this case convincing, but perhaps there's a way to undermine it. My second line of reply, then, is this. Whether or not I can construct a thought-experiment to specification, it seems to me that it's in the nature of representation (at least of the extra-mental) that Frege cases are possible. I don't see how you can have a representational system that relates the mind to external objects and properties, and for which the possibility of type-distinct representations referring to the same item, giving rise to ignorance of the identity, does not exist. So in a way I want to shift the burden of argument. Rather than focus on the details of any alleged sensory Frege case, I want to pose the following challenge to the advocate of ERQ: if indeed sensation is to be understood as a species of representation on a par with language (with some differences, admittedly), then what could there be in this case of mind-world relation that, in principle, prohibits the possibility of a Frege case? But if it's possible, and qualitative character is reduced to (external) representational content, then the potential for an incompatibility with self-knowledge of qualia exists. In the face of such incompatibility, it's clear to me that we should keep the self-knowledge and lose ERQ.

## Notes

- 1 See Dretske (1995), Lycan (1996) and Tye (1995).
- 2 For physiological type reductionism, see Hill (1991), and for functionalism see Levin (1986).
- 3 See Shoemaker (1981) and Block (1990).
- 4 See in particular Rey (1992).
- 5 I suppose one could add the position that representation is a brute relation. For obvious reasons this isn't really a live option, and anyway this is less a theory of representation than a denial that there is such a theory.
- 6 Fodor (1994) has abandoned narrow content, and Lycan (at least in conversation) deplors it. In its defence see Rey (1998).
- 7 See Burge (1988) and Gertler (2000).
- 8 As advocated by Fodor (1994).
- 9 I say only 'in a sense' because, as we saw above, there is a perfectly good sense in which I do know the identity of my thought that I'm drinking water – it's the thought that I'm drinking water. But what I can't know, merely from within, is whether I live on Earth or Twin Earth, and therefore whether my thoughts are about H<sub>2</sub>O or XYZ.
- 10 There's a complexity I need to address here. On some theories of thought individuation, such as Salmon (1986), the thought that I'm drinking water just is the same as the thought that I'm drinking H<sub>2</sub>O. This is an extreme externalism, and I don't find it plausible, but it's not crucial to my point to deny it. On this extreme externalism, then, it wouldn't be true that one could tell merely from within that two thoughts were of distinct types. But surely there is some mental difference between the two states, and if it's not to be marked as a

- difference in thought type, then call it something else. The point is, there is a distinction concerning my state of mind that I can make and that isn't hostage to the external facts.
- 11 How could this be so? I will address this question in more detail below. The idea is that since how things seem is a function both of how the world is and how I am, the difference in experience could be attributable to me.
  - 12 Most recently, Levine (2003).

## References

- Block, N. (1990), 'Inverted Earth', in J. Tomberlin (ed.), *Philosophical Perspectives, 4: Action Theory and Philosophy of Mind*, Atascadero, California: Ridgeview Publishing, 53–80.
- Burge, T. (1988), 'Individualism and Self-Knowledge', *Journal of Philosophy*, 85, 649–63.
- Dretske, F. (1995), *Naturalizing the Mind*, Cambridge, MA: MIT Press, Bradford Books.
- Fodor, J.A. (1994), *The Elm and the Expert: Mentalese and its Semantics*, Cambridge, MA: MIT Press, Bradford Books.
- Gertler, B. (2000), 'The Mechanics of Self-Knowledge', *Philosophical Topics*, 28, 125–46.
- Hill, C.S. (1991), *Sensations: A Defense of Type Materialism*, Cambridge: Cambridge University Press.
- Levin, J. (1986), 'Could Love Be Like a Heatwave? Physicalism and the Subjective Character of Experience', *Philosophical Studies*, 49, 245–61.
- Levine, J. (2003), 'Experience and Representation', in A. Jokic and Q. Smith (eds), *Consciousness: New Philosophical Perspectives*, Oxford: Oxford University Press.
- Lycan, W.G. (1996), *Consciousness and Experience*, Cambridge, MA: MIT Press, Bradford Books.
- Rey, G. (1992), 'Sensational Sentences Switched', *Philosophical Studies*, 68, 289–319.
- Rey, G. (1998), 'A Narrow Representationalist Account of Qualitative Experience', in J. Tomberlin (ed.), *Philosophical Perspectives 12: Language, Mind, and Ontology*, Oxford: Blackwell, 435–57.
- Salmon, N. (1986), *Frege's Puzzle*, Cambridge, MA: MIT Press.
- Shoemaker, S. (1981), 'The Inverted Spectrum', *Journal of Philosophy*, 74, 357–81.
- Tye, M. (1995), *Ten Problems of Consciousness*, Cambridge, MA: MIT Press, Bradford Books.