**Expressivism and the Metaphysics of Consciousness**

**Nicholas Unwin**

**Abstract**

An expressivist theory of consciousness is outlined and defended. The suggestion that attributions of consciousness to others involve an essentially projective element is argued to be implicit in many influential views about the mind. Creation stories, as standardly used to illustrate supervenience claims, are shown to be helpful. The view that ‘zombism’, defined as the disposition to think that other people are unconscious although physically normal, is not a narrowly cognitive disorder is defended, and a comparison is drawn with the Capgras delusion. It is argued that consciousness and values behave similarly with respect to weak and strong supervenience on ordinary physical facts, though empathic projection does not easily combine with modal distance. Chalmers’ objection that a projectivist stance makes no sense with regard to one’s own consciousness is considered and rejected.

**Keywords**: expressivism; projectivism; consciousness; zombie; Capgras delusion; modality; supervenience

1 Introduction

Expressivist theories have long been influential in ethics. They are useful because they promise to explain why moral values seem to be something over and above the purely factual considerations that underpin them, and yet without importing mysterious new qualities. It has often been noted that consciousness is not unlike values in the sense that it too intuitively seems to involve something over and above purely physical phenomena; and yet, at the same time, any new non-physical properties that might make up the difference will inevitably turn out to be unacceptably mysterious. There is an obvious similarity here, and yet there have been no serious attempts to develop an expressivist theory of consciousness. It is not that the possibility has never been considered. It is just that it is generally assumed that such a project will fail, and for fairly straightforward reasons.[[1]](#footnote-1) This paper will argue that the connection between consciousness and values deserves to be taken more seriously; and, although I shall not conclude that expressivism definitely works here, success is sufficiently close to be highly instructive.

2 Consciousness as something projected

What is expressivism? The core idea, as applied to ethics, is that a person’s moral judgements do not express only beliefs about what is the case, but also a wider range of mental states, such as feelings and desires. It follows (given a few additional assumptions) that the sentences which our moral judgements express are not wholly declarative, even though they certainly have that appearance. Thus, a major task faced by traditional expressivists is to give an alternative account of ethics that explains how such non-declarative sentences can adequately do the job required of them.[[2]](#footnote-2) However, there are also weaker views which are expressivist in tone but which arguably do not require such revisionist semantics.

Do our thoughts about consciousness also include a noncognitive component? The idea may sound odd, but there are some influential ideas that seem to move in this direction. There is, for example, the famous remark of Wittgenstein’s (1958 IIi): ‘My attitude towards him is an attitude towards a soul; I am not of the opinion that he has a soul.’ If I see someone in pain, I do not merely see, in a detached sort of way, a human body exhibiting pain-behaviour. On the contrary, I interact, empathize and so forth, and these attitudes form a significant part of my recognition that this person is in pain. This gets close to saying that to judge that a person is in pain just *is* to express certain noncognitive (as well, of course, as cognitive) attitudes of a certain kind. At any rate, the distinction between opinions and attitudes is clearly expressivist in tone.[[3]](#footnote-3)

Another candidate for consideration is Dennett (1989), who insists against Fodor, for example, that intentional interpretations of people’s behaviour are projections from what he calls an ‘intentional stance’. Intentional facts have no reality outside the stance (there are no ‘unmeant meaners’ in the way in which Aristotle thought there had to be an unmoved mover). Although Dennett disavows expressivism, the connection is very clear. Intentional facts are not part of the fabric of the universe in the way in which physical facts are. Rather, they are projections of a certain kind.[[4]](#footnote-4)

Thirdly, and from a very different tradition, we have Collingwood (1983, ‘Epilegomena’), who insists that to understand another’s action we cannot just look at the ‘outside’ of the event. We have to grasp the ‘inside’ as well, and the only way to do that is to ‘re-enact’ the thought behind it in one’s own mind. He is concerned primarily with thoughts rather than the desires and feelings that might also underlie actions, but the idea generalizes naturally. On such a view, to understand another’s thoughts and feelings is to experience them oneself, albeit in a detached, attenuated and modified form. Modern ‘simulationist’ accounts of how to interpret other people’s mental states adopt a similar view, and there is clearly an expressivist or projectivist element here.[[5]](#footnote-5)

How would this help us with the metaphysics of consciousness? The chief problem here is that neither physicalism nor dualism seems satisfactory, although they appear to exhaust all the alternatives. Physicalism identifies consciousness with physical states, and as a consequence demands that it be impossible to have a creature with all the right physical characteristics (the characteristics that you yourself have, for example) and yet not be conscious: in short, that consciousness supervenes on the physical. This is to ignore the powerful Cartesian intuition that ‘zombies’ (i.e. unconscious duplicates of ourselves) are, at least, logically possible (though presumably not naturally or causally possible).[[6]](#footnote-6) Yet dualism imports qualities that are essentially mysterious—weird, ghostly entities and properties that do not fit into any normal scientific picture of the world. If we could find a third position that accounts for the logical possibility of zombies, and yet which does not import anything more into the universe than the physicalists do, then we shall have evidently achieved something extremely useful: and some sort of projectivism looks as if it might fit the bill exactly.

3 Projections and supervenience

Let us look at this in more detail. Our initial problem is one of supervenience, and this is an idea that is often illustrated using creation stories. Thus, on the First Day, God creates light; and (in my version, at any rate) the complete distribution of elementary particles as well, together with any other microphysical phenomena there might be. On the Second Day, He turns His attention to macroscopical things—mountains, rocks and fjords, for example—but discovers that they have already been taken care of. Any alteration to the latter can only be achieved by revising the previous day’s work. This is what we mean when we say that mountains, rocks and fjords (and so forth) are *supervenient* on the microphysical world: it is impossible to have two situations alike in respect of the latter, but unalike in respect of the former. God therefore considers what else needs doing, and decides to create life. The trouble is that, once again, He discovers that He has already arranged that as well. He has certainly created DNA molecules, plants and animals, such as human beings, and with (it would seem) all their normal biological functions. He considers introducing some *élan vital* to add a little zest to things but, upon reflection, decides that there would be no point. Life, too, is clearly supervenient on the microphysical. Filled with forebodings as to the consequences of bequeathing a one day working week to posterity, He nevertheless accepts the inevitable and decides to take early retirement.

But then it is brought to His attention that there might actually be two things needed that have not yet been created: consciousness and values.[[7]](#footnote-7) Logically, it would make more sense to create consciousness first, for who needs values if there is no pain? But let us look at values first anyway. God has to consider what else needs to be done to infuse His creation with values, and it is here that He is forced to notice that objective values, in Mackie’s (1990) phrase, would have to be very ‘queer’ things indeed. Yet it is also clear that you cannot get an ‘ought’ from an ‘is’, so it cannot be that values were automatically created on the First Day: there is something which still needs to be done even after all the ‘facts’ are in place. Some sort of expressivism seems to be called for; and, with that in mind, we naturally ask how God might come into it. Well, what He actually did, as a matter of historical record, was to issue some commandments (eventually); but the interesting point is that these imperatival acts in no way enriched what Mackie calls ‘the fabric of the world’, any more than human prescriptions do. In that sense, He did not actually *add* anything at all to His creation: certainly not ‘objective values’, whatever they are. But to say that He therefore *did* nothing at all that is relevant would also be wrong. Among other things, He legislated, and that is certainly to do something. But laws are not quite the same as moral norms, and are metaphysically not nearly as puzzling. The introduction of ethics into the world is less easily understood.

What actually happened on the Second Day (at least, according to my own revisionist theology) is that, although the world itself did not change, God did. Having hitherto adopted a morally indifferent attitude towards His creation, He then started to engage with it in a way which essentially involved the affective and conative parts of His mind. The use of temporal language here is potentially misleading, of course, for He cannot suppose Himself to have *created* values by virtue of His making these prescriptive judgements, if that is to imply that murder was not wrong before He forbade it, or would not have been wrong had He not forbidden it: that is to make the standard mistake of confusing prescriptivism with moral subjectivism.[[8]](#footnote-8) But if God Himself is to recognize that the values that His creatures are talking about are genuinely real, then He would have to share their attitudes (or at least some of them). And His recognition of their reality would be actually constituted (at least, in part) by His sharing these attitudes; which is clearly an expressivist position.

With consciousness we get an analogous situation. Cartesian souls, like objective values, are ‘queer’ entities, and it is unclear that the zombies created on the First Day will become conscious just by adding more furniture to the world. The zombies themselves will, of course, insist indignantly that they are already conscious, and the physicalists among them will likewise insist that attempts to construct a ‘body–mind’ gap by analogy with the ‘is–ought’ gap are quite uncalled for. Nevertheless, even the most militant physicalist should concede that zombies at least *appear* to be conceivable (and therefore, perhaps, possible), and it would be useful to find a way of explaining this; and the crucial point is that there does seem to be an important difference between *judging* that a creature is a conscious human being and merely judging that that creature has all the purely physical characteristics of a human being: the latter judgement does not logically demand the former. The move from

What is it for *x* to be ϕ?

to

What is to for *S* to judge that *x* is ϕ?

is, of course, the first step in expressivist analyses. With this in mind, we are tempted to say, once again, that what needs to happen is for God to change Himself rather than His creation. That is to say, to introduce consciousness into the world (or, more accurately, His view of the world), He needs to stop looking at people as if they were merely things, but instead interact and empathize, adopt the intentional stance, simulate or re-enact some of what this stance reveals, and so on. Like Pygmalion, He might even start to love His creatures.

The immediate objection is that such attitudinal changes cannot possibly create actual consciousness. Pygmalion may have fallen in love with the statue he created, but the statue was not thereby rendered conscious. That had to wait until Venus took pity on him and brought it to life, a process that presumably meant actually doing things to the statue, and not just thinking about it in a different way. Likewise, no amount of empathizing, re-enacting, adopting the intentional stance, or any other kind of attitude-towards-a-soul type of activity will turn unconsciousness into consciousness. Yet we have the same intuitions about expressivist ethics. I cannot seriously suppose that things are made genuinely good or bad just because I (or anyone else) develop certain attitudes towards them, it will be insisted; and, of course, this is perfectly right. However, expressivist theories do not require us to suppose that values are actually *constituted* by our attitudes in this sense. What expressivists do argue to be impossible, rather, is for *S* to have certain non-cognitive attitudes towards *x* without *S* *herself* thereby judging that *x* is wrong (and vice versa). Similarly, we might wonder whether it is possible for Pygmalion to fall (genuinely, non-pathologically) in love with his creation without also himself judging his creature to be alive and conscious (we do not have to share his judgement). A negative answer here is not so absurd.

What is crucial is that the First Day’s creative work was not, in any immediate sense, factually deficient. Although no mention was made then about consciousness or values, that does not mean that the world’s furniture was not all yet in place. But at the same time, there is at least a sense in which it is true that consciousness and values were not yet firmly in place, and that more work needed to be done. Articulating what else is needed is hard, and many will find my story about God’s projections to be bizarre or worse. But we can expect some tension here, for consciousness is about subjective experiences and how things are seen from a particular point of view, whereas what philosophers call the ‘God’s eye point of view’ is typically expected to be wholly objective, neutral and detached. So to ask what consciousness is ‘in itself’ or ‘objectively speaking’ is bound to lead to difficulties, difficulties which these creation stories manage to illustrate reasonably well.

4 What is it to judge that *x* is a zombie?

We need to give an account of a strange condition, which I shall call ‘zombism’. Normally, a zombist is understood to be someone who merely thinks that zombies are possible: she is thus not required to think that any human being in this world is actually a zombie, only that some possible worlds might contain such creatures. We shall see, however, that modal judgements of this kind run into difficulties, and shall therefore consider, in the first instance, a much stronger condition. A ‘zombist’, on my definition, is someone who thinks that all other people actually are zombies, or at the very least, refuses to attribute consciousness to any other humans. In other words, a zombist is a kind of extreme sceptic about other minds. Obviously, there is something very badly wrong with such zombists, but the question is what type of thing is wrong. On the standard view, they go wrong because they refuse to accept various basic facts about the world: the attitude they refuse to express is thus an example what I shall call a *narrowly cognitive* attitude. On my view, by contrast, the narrowly factual beliefs could all be in place: it is attitudes of a different kind that are missing.

What sort of attitudes are they? If I judge that X is conscious, then I judge that there is something which it is like to be X. The attitude expressed here is, I suggest, not narrowly cognitive, but rather invokes wider mental capacities such as imagination and empathy. There are probably also affective elements as well. It is highly controversial just where the boundary between the cognitive and the noncognitive should be drawn, but the boundary between what is narrowly cognitive and what is not is a bit more straightforward, and this is the distinction that I shall mostly rely on. Moral expressivists likewise insist that the attitudes expressed in moral judgements are noncognitive, or at the very least have a component which is not narrowly cognitive. The ethical analogue of a zombist is an amoralist, i.e. someone who refuses to accept any moral judgements, and she need not be wrong about any straightforward facts about the world. Everything on the ‘is’ side of the ‘is–ought’ gap could well be in place. This is not to belittle the seriousness of the condition: a consistent amoralist would be so odd as to be virtually unimaginable. The reason is that the attitudes that are malfunctioning here play a central role in our lives, and should therefore not be ignored or downplayed.

The distinctions between different kinds of attitudes are, to repeat, not well understood, and this ensures that the exact content of many of our judgements is unclear. To say that a person is *sighted*, for example, is, among other things, to tell a purely physical story about how a person responds to light, a story that could also apply to her zombie twin: after all, there is no obvious reason why a zombie should not score full marks when she goes to the optician for an eye test. Yet an unconscious person cannot see anything, so we need to say what it is that a zombie actually can do instead, and this is hard. The problem is that consciousness *per se* is not a single uniform quality common to conscious states of all kinds, something that can be readily peeled off mental functioning to leave a zombified residue. This is a fact that has sometimes been used to show that the very idea of a zombie is impossible.[[9]](#footnote-9) Most of our psychological concepts are mentally ‘thick’ in a way that makes being a zombie a condition that is hard to describe coherently.

However, the situation seems to be no worse here than it is in ethics, expressivism’s original home. Indeed, there is a striking parallel. For example, a central theme in ethics concerns the distinction between ethically ‘thin’ concepts such as *ought*, *good* and *right*, and ethically ‘thick’ concepts such as *courageous*, *cruel* and *generous*. The latter include factual elements that plainly relate to narrow cognition, not affect or conation. For example, to call someone courageous is, among other things, to tell a factual story about how she reacts to perceived danger. Many moral philosophers, notably G.E.M. Anscombe (1958) and Bernard Williams (1985), insist that thick concepts do most of our ethical work, and cannot be decomposed into a thin, purely ethical component and an ethically neutral, purely factual residue. Indeed, the very distinction between the factual and the evaluative, along with the ‘is–ought’ gap, is called into question. If this is right, then expressivists will clearly have difficulties. But the difficulties are perhaps surmountable.[[10]](#footnote-10) At any rate, my central thesis seems to be confirmed, namely that the metaphysics of consciousness really does resemble the metaphysics of values.

But can we really make sense of zombism as a psychological condition? The idea needs fleshing out a bit, and we may note that there certainly are some (actual) disorders that seem to resemble zombism, many of which are routinely discussed in the philosophy of mind. For example, autism is often characterized as a kind of ‘mind-blindness’, or inability to develop a sense that there are other minds. However, the resemblance is weak. Even severely autistic people recognize that other people feel pain (I think). The problem, rather, is with intentional states, such as beliefs and desires, not *qualia*. Moreover (and crucially), it is not just that autistic people cannot easily empathize with others; they have great difficulty in predicting other people’s (overt, physical) behaviour as well. This, evidently, is not the kind of condition we need to be interested in here.

Anyone who can successfully predict other people’s behaviour, but who nevertheless regards them as mere unconscious automata, sounds unpleasantly psychopathic, and we might therefore suspect that it is not autism but some kind of antisocial personality disorder that zombism most closely resembles. The connection between zombism and amoralism is certainly brought to the fore, once again. Yet, for all that, the resemblance between psychopathy and zombism is only moderately convincing, if only because zombism, unlike psychopathy, demands beliefs and attitudes that are wildly different from normal. Mildly psychopathic people, by contrast, are often not diagnosed as such, and have no difficulty in embedding themselves in normal social life, and it is unclear what a mildly zombist person would be like.

A slightly better candidate, perhaps, is the rare but much discussed condition, the Capgras delusion, where sufferers develop the strong conviction that someone close to them (a husband or wife, for example) has been replaced by an impostor. The sufferer is unable to specify any actual, observable difference between the original and the impostor on which this judgement might be based, which is philosophically baffling in itself, and in a clearly relevant way given that our attitudes apparently need to supervene on our beliefs.[[11]](#footnote-11) It seems that there are two mechanisms involved in the recognition of familiar people, one of them purely sensory and the other largely affective: and what happens with Capgras patients is that the mechanisms become dissociated. However, all known patients, unlike our hypothetical zombists, tend to suffer from either advanced schizophrenia, dementia or brain damage; so, again, the connection is not all that strong as it stands. Our zombists are meant to be odd, but not seriously psychotic (at least not in their narrowly cognitive lives). However, the feeling of unfamiliarity, which is what concerns us here, is presumably detachable, at least in principle, from the seriously pathological and narrowly cognitive error (namely, the belief that people have been replaced by impostors). So the residual disorder, should it ever exist, might be very similar to zombism, even though mere unfamiliars fall some way short of being full-fledged zombies. No doubt a trawl of the psychiatric literature could yield a few more interesting examples.

5 Weak and strong supervenience

However, we are not primarily concerned with whether human beings around us are zombies, but rather with whether some of the inhabitants of certain possible worlds are. To understand the effects of modalization, we evidently need to look more carefully at the notion of supervenience, together with some associated arguments that have seemed to many to prove that consciousness and values are quite different in this respect, and that attempts to extend projectivist analyses from the latter to the former are thus doomed to fail. Let us begin with the standard distinction between weak and strong supervenience:

*A*-properties *weakly supervene* on *B*-properties if and only if for any possible world *w* and any individuals *x* and *y* in *w*, if *x* and *y* are *B*-indiscernible in *w*, then they are *A*-indiscernible in *w*.

*A*-properties *strongly supervene* on *B*-properties if and only if for any possible worlds *w*1 and *w*2 and any individuals *x* in *w*1 and *y* in *w*2, if *x* in *w*1 is *B*-indiscernible from *y* in *w*2, then *x* in *w*1 is *A*-indiscernible from *y* in *w*2.[[12]](#footnote-12)

The difference concerns whether the two individuals in question inhabit the same possible world or (perhaps) different ones. Evidently, strong supervenience implies weak supervenience; however, the converse is false. One of Blackburn’s main arguments for projectivism is that values seem to supervene on ordinary facts only weakly, not strongly. This stems from the fact that if we judge *x* to be good it must be because we also judge that *x* has certain factual, good-making features. It therefore seems absurd for anyone to say of two entities *x* and *y* that one is good and the other not—even though they are indiscernible in all factual respects (or believed to be). Anyone who made such an invidious evaluative judgement would have to be so odd that we might not unreasonably doubt whether he was making a genuine evaluation at all, moral or otherwise. Hence, weak supervenience. Yet strong supervenience implies that if *x* is good, then in any possible world where *x* has the same factual properties as it does here and now, *x* is also good. This yields a modalized quantified conditional of the form

(SS) Necessarily: ∀*x* (*x* has total factual property F\* → *x* is good)

We thus appear to get an inference from fact to value, and this seems unfounded: as Blackburn puts it, one does not need to be logically confused (or factually inaccurate) to be a bad moralizer. We thus apparently have weak but not strong supervenience, and this yields a ‘ban on mixed worlds’, i.e., worlds where some entities are good and others not, despite being factually indiscernible; and Blackburn’s (1984, 1985) view is that moral realists will find this selective ban very hard to explain.[[13]](#footnote-13)

Do we have an analogous situation with regard to consciousness? It may be feared not. If we are allowed to suppose that zombies are possible at all (so consciousness fails to supervene strongly on physical properties), then there is no obvious reason why we should not have a conscious person and his zombie twin sitting next to each other, and hence a failure of weak supervenience as well. A mixed world seems to be no odder than any other, and Cartesian mental realism remains unscathed. True, it would be very odd to suppose, of two physically indistinguishable people right in front of you, that one of them is conscious and the other not. But it may nevertheless be insisted that, unlike the evaluative case, you would not automatically be guilty of a *conceptual* confusion if you were to suppose this. So is it really any odder than if they were put into different possible worlds?

Well, it is actually quite a bit odder. Imagine, if you will, a sufferer from what we might call the super-Capgras delusion. She is convinced that one of her identical twin sons has been replaced by an impostor, but not the other. Not only can she not say how the impostor fails to resemble the original he replaced; she cannot say how he fails to resemble his twin who has not been so replaced, even though they are both right in front of her. This hypothetical condition is surely a lot more extraordinary than its predecessor. We may still insist that a super-Capgras sufferer does not have to be conceptually confused, though since the concepts of personhood and impostorhood can be argued to have direct evaluative implications of their own, this is uncertain. But at any rate, mixed worlds clearly present more of a problem here than they might seem. The same point applies to zombism. On my original definition, a zombist rejects all other minds, and this is certainly a strange condition. But if instead we have a ‘super-zombist’, who attributes mentality to one person but withholds it from his identical twin, then we have a pathology that is considerably harder to make sense of, since there can be no *grounds* of any kind on which this selective attribution can be made. Some sort of conceptual error (amongst other things) is thus surely involved here, contrary to what was suggested above. The analogy between consciousness and values therefore seems to hold up quite well when it comes to weak supervenience.

Strong supervenience is more problematic, and Blackburn is surely wrong in two respects. Firstly, our intuitions about supervenience extend to cross-world comparisons. If I think that *x* is good because it has good-making characteristics F\*, then I shall naturally insist that *if* *x* were to have F\*, then it *would* be good. In other words, in any possible world where it has F\*, then it is good, and this yields (SS). Secondly, (SS) will not yield an inference from the factual antecedent to the evaluative consequent unless it is also *a priori*, yet the modality is meant to be metaphysical, not epistemological.[[14]](#footnote-14) This perhaps is answerable by noting that a person who evaluates wrongly despite having no false nonmoral beliefs does not even seem to be in metaphysical error, let alone a contradiction, but people’s intuitions may vary here. Moreover, the basic Humean point, that we cannot get from an ‘is’ to an ‘ought’, or from factual premises to an evaluative conclusion, has its origin in the separateness of beliefs and passions: factual statements express the former, whereas value judgements express the latter. This is why, on the expressivist view, value judgements are not ordinary declarative sentences, but utterances like ‘Hooray to *x*!’ (usually abbreviated as ‘*H*!*x*’). But this ensures that (SS) needs to be formulated as

(SS′) Necessarily: ∀*x* (*x* has total factual property F\* → *H*!*x*)

and it is unclear how to understand this formula. The ‘Frege-Geach problem’ of explaining how to attach meanings to embedded evaluations, and in such a way as to preserve ordinary inferences, is still unsolved even when dealing with simple sentential connectives. Here, ‘*H*!*x*’ falls within the scope of, among other things, a modal operator, and I know of no attempted solution to the Frege-Geach problem that addresses modal contexts. (SS′) looks like, among other things, a kind of complex evaluation, perhaps the expression of praising anything F\* under any circumstances (even in possible worlds where our tastes are different); though it could also perhaps be understood as expressing a rejection of any metaphysical ‘is-ought’ gap. Since we seem to have two rather different ideas here, this complicates the picture considerably. At any rate, we can see why empathic projection and modal distance make a hazardous combination.

Thus in both moral and mental cases we have weak supervenience, and there is no straightforward inference either from the factual to the evaluative or from the physical to the conscious. There is more of a question mark over strong supervenience. Obviously, much more needs to be said on how best to understand both moral and mental supervenience claims, but for present purposes it is sufficient to note that, once again, the problems seem to run parallel.

6 Projectivism and the immediacy of consciousness

We must now consider another, and perhaps even more basic line of objection to a projectivist theory of consciousness, namely that consciousness has an immediate and undeniable quality which moral values do not have. Thus David Chalmers (1996: 83) writes:

… moral facts are not phenomena that force themselves on us. Indeed, this reflects the strategy taken by moral antirealists such as Blackburn and Hare. These antirealists argue that because moral facts are not entailed by natural facts and not plausibly ‘queer’ further facts, they have no objective existence and morality should be relativized into a construct or projection of our cognitive apparatus. The same strategy cannot be taken for phenomenal properties, whose existence is forced upon us.

But is this right? I have so far been concerned only with other minds, not one’s own experience, and it is unclear that we have a sharp difference between values and other minds in respect of what is forced on us. Hume famously wrote that ‘’[t]is vain to ask whether there be body or not. That is a point we must take for granted in all our reasonings’. In other words, we just have to believe in ‘bodies’ whether we like it or not. Exactly the same claim applies to other minds: unless we are seriously deranged we shall assume that other human beings have minds, just as we shall assume that there is a tree in the quad if given the right sort of sensory stimuli. These conclusions we find overwhelming. But it is also the case that, unless we are likewise deranged, we shall feel immediate moral outrage when confronted with certain phenomena. Although it is Nature, not Reason, which sustains these responses (according to Hume), negative claims about bodies, other minds or moral values with regard to ‘objective existence’ seem equally unintuitive. The reality of both moral values and other minds seems overpoweringly obvious, and in roughly equal measures.

However, it may be protested that the real problem concerns one’s own mind, not other people’s. That I myself am conscious is an immediate datum whose reality appears to be utterly fundamental. Moreover, this Cartesian assuredness is surely grounded in what Hume calls ‘Reason’ if anything is! Yet, although this might argue against certain kinds of anti-realism, it need not undermine every kind of projectivist position. All that my version requires here is that it be impossible for a person to *judge* that he himself is conscious without his thereby adopting a certain kind of non-cognitive empathic attitude towards himself, and vice versa. The move from the former to the latter is not obviously harmful in any way. The move back may seem odder, but in so far as attitudes require consciousness by definition, it may be no more than a trivial truth. We are not thereby required to conclude that our own consciousness ‘should be relativized into a construct or projection of our cognitive apparatus’, as Chalmers puts it. Consciousness remains utterly real even if the relevant discourse has slightly unusual features.

The idea may still sound very odd, as though we somehow ‘bootstrap’ consciousness into existence through adopting certain attitudes towards ourselves, an idea that seems to put the cart before the horse. Against this, two points should be remembered. Firstly, we have already seen that projectivism, unlike subjectivism, does not obviously require this: projected properties are not mind-dependent in the relevant way. For example, Simon Blackburn (1981: 174–5), in a well known passage, insists that

[i]f most of us come to taste phenol-thio-urea as bitter, then that is *what it is* for the stuff to become bitter. If most of us come to find wanton violence admirable, that is not what it is for wanton violence to become admirable: it is what it is for most of us to deteriorate in a familiar and fearful way.

Thus judgements about values and secondary qualities behave quite differently when put into temporal contexts. The same point applies to counterfactual contexts, so although phenol-thio-urea would have been bitter if I and everyone else had always tasted it that way, wanton violence would not have been admirable even if I and everyone else had always found it to be so: the possible world envisaged is simply one in which everyone is, and always has been, just wrong. What is happening is that I, from the outside, am projecting my own actual view about wanton violence into this possible world (including my depraved counterpart), which I am certainly allowed to do. It is in this sense that values are not mind-dependent at all, and the ‘realism’ in Blackburn’s term ‘*quasi*-realism’ becomes appropriate. Likewise, and for the same reason, it remains true that, given that I *am* conscious, I would *still* have been conscious even if everyone (including myself) had failed to adopt certain consciousness-attributing attitudes towards me. At least, that is true on one way of reading it. The basic reality of consciousness, to that extent, is just not threatened by my projectivist story (quite the reverse), and the cart has not been put before the horse. To put it slightly paradoxically, consciousness is not a mind-dependent quality! There is much more to be said here, of course, and the paradoxes need further unravelling. But it suffices to note that my theory is not so ambitious that it claims to make consciousness into something wholly unmysterious; it merely aims to help us to see where some of the mysteries come from.

Secondly, we should remember that ordinary physical features will still be in place regardless of how or whether we project anything onto ourselves. In particular, our brain processes will be in place. This will not impress dualists such as Chalmers, of course, but physicalists (i.e. most philosophers nowadays) suppose that these features are themselves sufficient to guarantee consciousness, and can therefore hardly go on to claim that these features taken together with the projections will *not* guarantee consciousness! The complaint, rather, would have to be that the alleged ‘bootstrapping’ is unnecessary, not that it is insufficient. It is worth stressing this point, since although my theory may seem to many to be grossly implausible, it is very close to a kind of amplified physicalism which is not at all implausible. This view insists that zombies are indeed metaphysically impossible, but nevertheless offers an account of why they appear otherwise. The point is that when we judge a person to have ordinary mental qualities, we tend to invoke a much wider range of attitudes than those required to attribute basic physical qualities. This can easily lead us to think (wrongly, it is held) that there must be more to the mental than just the physical. Such a view should attract physicalists, since it is always useful if your theory includes an explanation of why it seems to its opponents to be wrong. Yet it is also clearly very close to my theory, and so the latter cannot really be as far-fetched as it might seem at first sight. Moreover, this amplified physicalism does have its disadvantages when compared with my view. If even basic judgements about other minds involve attitudes that are not expressed when we make judgements about the supposed physical underpinnings, then it is hard to see how minds can really be nothing over and above these physical underpinnings. If they were, then the additional attitudes should drop out of the system as irrelevant, and they do not seem to do so.

In any event, the idea that psychological self-ascriptions have an inbuilt expressive element is hardly new. Wittgenstein’s avowal theory is the obvious example here, and the idea has been developed, for example, by Dorit Bar-On (2004). These anti-Cartesian theories are primarily designed to explain why I can seldom be mistaken about what is going on in my mind right now, rather than what self-awareness actually is, but they could perhaps be extended in that direction.

So is this thesis about consciousness correct? I have certainly not proved that it is; but, significantly, I have nevertheless shown that it is far less implausible than it is usually taken to be. Our intuitions about whether zombies are possible are clearly unstable: the unending disagreement is enough to prove that. Projectivism could explain why that is, why supervenience claims are inevitably going to be problematic here. The suggestion that ‘zombism’, should it ever exist, would be primarily an attitude problem rather than a narrow-belief problem is not at all unreasonable. And projectivist theories, if they can be made to work, are metaphysically modest, which is just what we need when dealing with the mind–body problem.

**Bibliography**

Anscombe, G.E.M., ‘Modern Moral Philosophy’, Philosophy **33** (1958), pp. 1–19

Bar-On, Dorit (2004), *Speaking My Mind: Expression and Self-Knowledge*. Oxford: Clarendon Press

Blackburn, Simon (1981), ‘Reply: Rule-Following and Moral Realism’, in Steven Holtzman and Christopher Leich (eds), *Wittgenstein: To Follow a Rule*. London: Routledge & Kegan Paul: 163–87

———(1984), *Spreading the Word: Groundings in the Philosophy of Language.* Oxford: Clarendon Press

———(1985), ‘Supervenience Revisited’, in J. Casey (ed.) *Morality and Moral Reasoning*. London: Methuen

———(1990). ‘[Hume and Thick Connexions’.](http://www.jstor.org/sici?sici=0031-8205(199023)50%3c237:HATC%3e2.0.CO;2-I) *Philosophy and Phenomenological Research* **50**: 237-250

———(1993), ‘Moral Realism’, in *Essays in Quasi-Realism*. Oxford: Oxford University Press: 111–29. Originally printed in Casey (ed.) above.

Chalmers, David (1996), *The Conscious Mind: In Search of a Fundamental Theory*. Oxford: Oxford University Press

Collingwood, R.G. (1983), *The Idea of History.* Oxford: Oxford University Press

Dennett, Daniel C. (1989), *The Intentional Stance*. Cambridge MA: MIT Press

Gibbard, Allan (1990), *Wise Thoughts, Apt Feelings.* Oxford: Clarendon Press

Kim, Jaegwon (1993), *Supervenience and Mind: Selected Philosophical Essays.* Cambridge: Cambridge University Press

Little, Margaret Olivia, ‘Virtue as Knowledge: Objections from the Philosophy of Mind’, *Noûs* **31** (l997): 59–79

Mackie, J.L. (1990), *Ethics: Inventing Right and Wrong*. London: Penguin

McLaughlin, Brian and Bennett, Karen, ‘Supervenience’, *The Stanford Encyclopedia of Philosophy (Summer 2010 Edition)*, Edward N. Zalta (ed.), URL = <http://plato.stanford.edu/archives/sum2010/entries/supervenience/>

Nagel, Thomas (1979), ‘Panpsychism’, in *Mortal Questions*. Cambridge: Cambridge University Press: 181–95

Ratcliffe, Matthew (2007), ‘What Is a Feeling of Unfamiliarity?’, *Philosophy, Psychiatry, & Psychology* **14.1**: 43–9

———(2008) ‘The phenomenal role of affect in the Capgras delusion’, *Continental Philosophy Review* **41**: 195-216

Rey, Georges (1995), ‘Towards a Projectivist Account of Conscious Experience’, in Thomas Metzinger (ed.), *Conscious Experience*. Paderborn: Ferdinand Schöningh: 123–44

Savellos, Elias E., and Ümit D. Yalçin (eds) (1995), *Supervenience: New Essays*. Cambridge: Cambridge University Press

Unwin, Nicholas (2001), ‘Norms and Negation: A Problem for Gibbard’s Logic’. The Philosophical Quarterly **51** (2001): 60–75

Wilkes, K.V. (1978), *Physicalism*. London: Routledge & Kegan Paul

Williams, Bernard, *Ethics and the Limits of Philosophy* (Cambridge MA: Harvard University Press, 1985)

Wittgenstein, Ludwig (1958), *Philosophical Investigations.* Oxford: Blackwell

Young, Garry (2007), ‘Clarifying “Familiarity”: Phenomenal Experiences in Prosopagnosia and the Capgras Delusion’, *Philosophy, Psychiatry, & Psychology* **14.1**: 29–37

Zangwill, Nick (1995), ‘Moral Supervenience’, *Midwest Studies in Philosophy* **20**: 240–62

1. See, for example, Chalmers 1996: 83. Nagel (1979) talks of ‘Realism’ (the view that mental properties are genuine properties), and recognizes that it is not obviously true: but concludes, nevertheless, that rejecting it makes little sense. He does not, however, explicitly consider an expressivist approach. [↑](#footnote-ref-1)
2. For example, Blackburn (1984) and Gibbard (1990). [↑](#footnote-ref-2)
3. Blackburn (1993: 112–3) considers Wittgenstein’s dictum here, though does not explore the expressivist implications in any detail. [↑](#footnote-ref-3)
4. *Projectivism* is best understood as a thesis about the status of certain entities or properties, whereas *expressivism* is best understood as a thesis about the associated discourse. However, since the theses fit together naturally, the terms are frequently used more or less interchangeably. This, I think, is harmless. [↑](#footnote-ref-4)
5. See also Ratcliffe 2008. [↑](#footnote-ref-5)
6. See Chalmers 1996, for example. [↑](#footnote-ref-6)
7. I am concerned primarily with the problem of other minds, and so will ignore, for the moment, the problem of God’s own consciousness. I shall also ignore the fact that God ‘saw that [His creation] was good’. [↑](#footnote-ref-7)
8. Legislation, by contrast, should not ordinarily be retrospective, let alone counterfactual. [↑](#footnote-ref-8)
9. See, for example, Wilkes 1978: 109–12. [↑](#footnote-ref-9)
10. Though not insuperable ones. For an account of this problem, and a suggested solution, see Blackburn 1990. [↑](#footnote-ref-10)
11. See, for example, Ratcliffe 2007, 2008 and Young 2007 for a discussion of some philosophical implications here. Georges Rey (1995: 137 n.27) also briefly discusses the Capgras delusion in a similar context. [↑](#footnote-ref-11)
12. See, for example, Kim 1993, Savellos & Yalçin 1995 and McLaughlin & Bennett 2010. [↑](#footnote-ref-12)
13. See also McLaughlin & Bennett 2010, §4.1. [↑](#footnote-ref-13)
14. See also Zangwill 1995. [↑](#footnote-ref-14)