

Expressivism and motivation internalism

Richard Joyce

[This is the penultimate draft of the article that appeared in *Analysis* 62 (2002) 336-344.]

Noncognitivism is the metaethical view according to which public moral judgements do not express beliefs, in spite of the fact that they are typically formed in the indicative mood. One form of noncognitivism — prescriptivism — holds that moral judgements are really commands.¹ Another form, on which we will focus — expressivism — holds that moral judgements function to express desires, emotions, or pro-/con-attitudes (in Simon Blackburn's words: 'a stance, or conative state or pressure on choice and action' [1993: 168]). Making a frequent appearance in the argumentative fray over noncognitivism is a thesis that is usually called 'motivation internalism':

MI: It is necessary and *a priori* that for any x and any y : if x judges that some available action y is morally right (good, obligatory, etc.), then x will have some (defeasible) motivation in favour of performing y .

The role of this thesis in the dialectic varies. Sometimes it is used as a premiss in favour of expressivism, as a desideratum that only expressivism can secure, or sometimes as a consequence of expressivism (and thus its denial implies cognitivism). It is, at the very least, usually assumed that expressivism and MI sit together comfortably. The task of this paper is to argue that expressivism neither implies, nor is implied by, MI. Of course, if either thesis is a necessary truth (as MI explicitly purports to be) then it will be strictly implied by the other. When I claim that expressivism does not imply MI (and vice versa), I mean that by endorsing expressivism one does not thereby commit oneself to the endorsement of MI (and vice versa) — something that is worth observing even if it turns out that MI (or expressivism) is implied, in some sense of the word, by every proposition.

The key to the argument that the affirmation of either thesis is consistent with the other's denial is a proper understanding of what it is for a kind of utterance to *express* a mental state. This understanding is a worthwhile end in its own right, quite apart from the issue of MI, since until we properly grasp this relation of *expression* we cannot hope to make headway on the problem of what kind of mental state moral judgements express — and yet this crucial issue has received remarkably little scrutiny.

1. Expressivism does not imply motivation internalism

Let us start the argument in an unexpected place, asking 'What is an apology?' There are two ways of taking the question. First, 'What are the criteria for an apology? — What does it take for an apology to have occurred?' Second, 'What is an apology for? — What, exactly, *is* an apology?' The first question is somewhat puzzling, but is hardly a philosophical mystery. We might say that in order for X to apologise to Y, there must be some understanding that Y has been wronged in a manner for which X is responsible (or X is an authorised representative of those who are responsible); X must say to Y the words 'I am sorry' (or some equivalent); X

¹ See, for example, Carnap 1935; Stevenson 1937. In several books and articles, R.M. Hare argued that although moral judgements are not actually commands, they do entail commands.

must say these words in a non-sarcastic tone of voice; Y must hear and understand these words; and so on. Now to the second and more interesting question: if these criteria are satisfied, and X succeeds in apologising to Y, what, exactly, has happened? Part of the answer is that X has *expressed his regret* to Y. (In addition, X has acknowledged the wrong done to Y and admitted responsibility for it.²) However — and this is the important bit — it doesn't follow that X *has* or *feels* the emotion in question. One can express regret — indeed one can express *one's own* regret — without actually having any regret at all. But how is it possible that X can express emotion E (at time *t*) without X's having E at *t*? The answer is simply that the notion of 'expression' that is in play does not denote a relation holding between X and his mental states, but rather denotes a more complex relation holding between X and Y, occurring against a background of established linguistic conventions in which X and Y are well-versed. These linguistic conventions decree that when the criteria outlined above are fulfilled, then, whether we like it or not (indeed, whether X likes it or not), X has, *inter alia*, expressed his regret to Y.³

This is quite different from the notion of 'expression' employed if we say that Emily's kicking over of her sand castle expresses her anger, or Chad's obsequious behaviour towards his ex-wife expresses his guilt over the failed marriage. In these latter cases, we really do require that Emily and Chad have the emotions in question, because part of what we mean is that these emotions cause or explain their actions. If it turns out that Emily doesn't have any anger, then her kicking over of the sand castle must have some other explanation — perhaps she was pretending to have anger, and pretending to express it via this unruly act. The person who apologises insincerely may also be interpreted as pretending, but he is not *pretending to apologise*, but is, rather, pretending to have regret, which is very different. Pretending to apologise, by contrast, is what an actor playing Sebastian in *Twelfth Night* will do during Act V.⁴

An insincere apology is a type of apology — unlike a fake Rembrandt, which is not a type of Rembrandt. This is more clear if we think of another case: promising. When one promises, one is, *inter alia*, expressing an intention, or commitment. If insincerity in some manner nullified the act of promising, then there would be no promise to break; but insincere promises *can* be broken just as can be sincere promises; therefore it is clear that insincere promises are still promises. (See Austin 1962.) Another case to think of is the making of an assertion, whereby one expresses one's belief. Lying is a species of assertion in which one expresses a belief that one does not in fact have. But again: a lie still succeeds in *being* an assertion. It is perfectly natural to say: 'George asserted that he has been to Italy, but it turned out that he was lying.' When George lies he is not pretending to assert, nor is he pretending to express a belief — rather, he is pretending to believe something. After all,

² Louis Kort (in his 1975) adds a couple more criteria: that the speaker is expressing regret for the offense as such, and that the speaker is performing an act of respect to his hearer as a person worthy of being spared mistreatment. See also Joyce 1999.

³ For a fuller account of these sorts of conventions — surrounding the act of promising rather than apologising — see Searle 1969: 57-61. See also Lewis 1969 for a cogent account of conventions.

⁴ I am not claiming that apologising while feeling no regret is *sufficient* for insincerity, for it seems that one could say 'Sorry' for a trivial misdemeanour, when distracted by some pressing matter (e.g., brushing someone while running to catch a bus), and nobody would go so far as to call it 'insincere,' but nor would anyone seriously claim that the speaker had the distinct emotion of *regret* at the moment of utterance. I return to this point below.

when we confront George with his mendacity, what we say is ‘You said you’d been to Italy, but you hadn’t, and you knew that you hadn’t — you lied!’ We *don’t* say: ‘You pretended to assert that you’d been to Italy, but you didn’t assert this at all — you lied!’ Of course, one *can* pretend to assert — actors do it all the time, but actors are not lying.

So: when one asserts *p*, necessarily one expresses one’s belief in *p*. From this one might be tempted to draw the following conclusion, which we could call ‘Belief Internalism’:

BI: It is necessary and *a priori* that for any *x* and any *y*: if *x* asserts *y*, then *x* believes *y*.

But we have seen that BI would be the *wrong* conclusion to draw; it is based on a misunderstanding of what it is to ‘express a belief.’ The denial of BI is perfectly compatible with the endorsement of the view that assertions necessarily express beliefs.

It is probably quite clear now where this is leading. There may be a kind of judgement that necessarily ‘expresses favour’ or ‘expresses a pro-attitude’ (however one prefers to word it), without it following that the speaker *has* the favour or pro-attitude in question. In other words, the expressivist may be correct about what moral judgements express, but MI would be the wrong way of capturing this truth. Indeed, one may admit all the counter-examples to MI (the listless agent, the evil agent, etc.⁵), thus accepting that it is not necessarily the case that moral judgement implies motivation, without this in any way undermining one’s confidence in the thesis that moral judgements express attitudes — attitudes that necessarily engage with, or are constitutive of, motivational states. This is an often-overlooked conclusion.

2. Motivation internalism does not imply expressivism

Sometimes it is thought that the acceptance of MI requires that one accept expressivism. Here are some thoughts (of nobody in particular) that are close to the expressivist’s heart:

To say that something is good is not merely to describe it, but is to evaluate it. To evaluate something positively is a way of expressing that you are in favour of that thing. To favour something is not to be indifferent towards it, but is to be disposed to pursue it, or promote it. To be disposed to pursue or promote something implies that one has some kind of desire-state regarding that thing. Desire-states necessarily engage (or perhaps are even constitutive of) motivations. Therefore, moral judgements necessarily involve motivations. But beliefs do not necessarily involve motivations, therefore moral judgements do not express beliefs; cognitivism is false.

The first person to argue along these lines (as far as I’m aware) was Hume:

Since morals, therefore, have an influence on the actions and affections, it follows, that they cannot be deriv’d from reason; and that because reason alone, as we have already prov’d, can never have any such influence. Morals excite passions, and produce or prevent actions. Reason of itself is utterly impotent in this particular. The rules of morality, therefore, are not conclusions of our reason. (*A Treatise of Human Nature* [1739] book 3, part 1, section 1)

⁵ See, for example, Stocker 1979; Milo 1981; Brink 1984, 1986.

Hume's argument relies on a second premiss (in addition to MI): that beliefs alone cannot provide motivation. The cognitivist might endorse MI but simply deny the premiss of Humean psychology. But even if the second premiss is granted the argument still doesn't work. From the fact that a certain kind of utterance is necessarily linked to a certain kind of mental state it doesn't follow that the utterance *functions to express* (in the sense discussed) that mental state. Consider again the act of apologising. The criteria for an apology involve a range of linguistic conventions in which both parties need be versed — for example, the addressee must hear and understand the words uttered, and the speaker must take it that this is the case. The satisfaction of these criteria will require both speaker and addressee to have certain *beliefs* — for example, the speaker must believe that his addressee hears and understands. This connection is a necessary (and *a priori*) one: it is not possible that any person could succeed in apologising to another person without having such a belief. And yet we would hardly say that the act of apologising — the utterance of 'I'm sorry' in the appropriate circumstances — *functions to express the belief that one's audience hears and understands*. Therefore, since a kind of speech act and a mental state may be necessarily linked without the former expressing the latter, MI does not imply expressivism.

3. Variations on motivation internalism and expressivism

It may be objected that I have misunderstood MI. Sometimes the thesis is presented with *sincerity* built into it:

MI_{sin}: It is necessary and *a priori* that for any *x* and any *y*: if *x* sincerely judges that some available action *y* is morally right (good, obligatory, etc.), then *x* will have some (defeasible) motivation in favour of performing *y*.

MI_{sin} does not imply expressivism, because if it did then so too would MI, since MI implies MI_{sin}, but, as we have seen in the preceding section, MI does not imply expressivism.

Does expressivism imply MI_{sin}? Here there is room for argument, but I believe that the answer is negative. Let us assume expressivism — that in making the judgement 'Stealing is wrong,' say, one thereby expresses disfavour towards stealing. (Nothing much should be read into my choice of the term 'disfavour' — it's standing in as proxy for whatever kind of mental state the expressivist wants to settle on — the important point is that it is a 'motivation-implicating' state.) One might think that from this follows something about what it would take for such a judgement to be *sincere*. For example, just as we might be tempted to think that a sincere apology is an expression of regret where one actually has regret, and a sincere act of thanking is an expression of gratitude where one actually has gratitude, and so on, then so too might we think that a sincere moral judgement is an expression of mental state *M* where one actually has *M*. And, under the expressivist assumption that the *M* in question is a motivation-implicating one, it follows that a sincere moral judgement must involve motivation.

However, this reasoning relies on a crude simplification of sincerity, not just regarding moral judgements, but also regarding apologies, thankings, etc. The argument turns on the following general claim, which (though perhaps seeming attractive at first glance) is false:

SINCERITY: S's utterance U (at time t) is sincere iff U functions to express mental state M, and S has M (at t)

Suppose Fred and Carol have to leave the dinner party unexpectedly, and as they rush out the door, dragging on their coats, they call out 'Thanks!' In fact, their minds were on some other matter, and at the moment of thanking they were not feeling any gratitude whatsoever. Nevertheless, we wouldn't ordinarily call their utterance *insincere* (though we may admit that it was less than heartfelt). Or suppose that just prior to their hurried departure, the dinner table discussion was on the morality of Britain's keeping the Elgin marbles. As they rush to the door, Fred (still distracted) gets in his two cents' worth: 'The marbles belong to the Greeks, and keeping them is wrong! Thanks and goodbye!' By ordinary lights Fred has certainly made a public moral judgement (albeit a hurried one) — and thus on our expressivist assumption he has expressed his disfavour towards the act of keeping the marbles — yet I do not think that the normal criteria for his judgement to count as *sincere* require that, at the moment of utterance, he had activated some particular kind of attitudinal/emotional state. If this is correct, then expressivism implies MI_{sin} only if we construe the appearance therein of 'sincerely' in some special theoretical way, different from ordinary usage.

One might well wonder what it matters if expressivism implies MI_{sin} (or vice versa), given that the latter appears to tell us something about only a proper subset of moral judgements. But I suspect that there is a train of thought according to which this appearance is deceptive — according to which moral judgements (unlike apologies, etc.) are *necessarily* sincere. On such a view, the appearance of 'sincerely' in MI_{sin} , while strictly redundant, may serve as a reminder of this attribute of moral judgements. Lying behind the thought that moral judgements are necessarily sincere is (presumably) the idea that moral judgements are really *mental* events, and only derivatively something that we do with language. Insincerity, in an apology or a promise, indicates a dissonance between what one is saying and what is going on in one's mind, but if a moral judgement is primarily just something that goes on in one's mind then the possibility of insincerity appears to retreat.⁶ In line with this mentalistic construal of the debate, we would accordingly revise the thesis of expressivism to be the view that moral judgements *are* certain kinds of desire, emotion, or pro-/con-attitude. And so the whole problematic *expression* relation disappears.⁷

Assuming that the explication of the expressivist's 'desires, emotions, or pro-/con-attitudes' turns out basically to mean 'motivation-implicating states,' then on this mentalistic construal of expressivism the connections between the theory and MI will be trivial. A moral judgement (where this is some kind of mental act) will necessarily be motivating; and from the fact that a kind of judgement necessarily implies motivation it will follow that such judgements can be considered motivation-implicating. Indeed, the connections will be so trivial that arguing for either thesis by means of first establishing the other ceases to be a

⁶ Complicating psychological phenomena like self-deception could potentially lead one to revise this claim.

⁷ According to this way of thinking, in the above example of Fred's hurried proclamation about the Elgin marbles, the expressivist would have to say that if at the moment of his departure Fred was not in some particular kind of emotive/conative state, then he did not *really* make a moral judgement at that moment. This in itself is, it seems to me, sufficiently counter-intuitive to cast the usefulness of the mentalistic version of expressivism into doubt.

feasible dialectical strategy.

In any case, taking the argument in this mentalistic direction makes the whole issue an empirical one — a very unwelcome result. According to such a view, we should, at least in principle, be able to take some persons who are paradigmatic instances of ‘moral judges,’ and with a PET scan watch for neural evidence of a ‘conative state’ — say, activity in the amygdala (or whatever) — when we prompt them to think about euthanasia, or Adolf Hitler, or returning the Elgin Marbles. But this, I think most will agree, is silly; nobody imagines that the cognitivist/expressivist debate can be settled in such a way.

The same unattractive conclusion follows from the view that when we claim that public moral judgements express mental state type so-and-so, we mean ‘express’ to denote a *causal* relation. But even this reasonably obvious observation — that the cognitivist/expressivist debate will not be settled with PET scans — pushes MI out of the debate, by forcing recognition of the fact that when we ask what a moral judgement is we are not investigating a kind of neural activity, but rather a linguistic activity; we are investigating not what *causes* these utterances, but their linguistic function. Let me end by drawing attention to the observation that once this fact is appreciated, the possibility clearly arises that moral judgements might express *more than one* type of mental state: both a belief and a desire, say.⁸ By contrast, thinking of the relation as causal might tempt one to assume that any such ‘joint judgement’ would have to flow from a kind of mental state with both belief-like and desire-like aspects — a ‘besire’, to use J.E.J. Altham’s term (see his 1986) — and an aversion to this psychology might lead one to reject the premiss. But the realisation that the relevant *expression* relation is non-causal allows us to see that no such conclusion follows. The fact that a kind of utterance may express belief *and* desire (say) implies nothing about the modal relations holding between the speaker’s belief-states and desire-states.⁹

References:

- Altham, J.E.J. 1986. The legacy of emotivism. In *Fact, Science and Morality*, eds. G. Macdonald and C. Wright, 275-288. Oxford: Basil Blackwell.
- Austin, J.L. 1962. *How to Do Things with Words*. Oxford: Oxford University Press.
- Blackburn, S. 1993. *Essays in Quasi-Realism*. Oxford: Oxford University Press.
- Brink, D. 1984. Moral realism and the skeptical arguments from disagreement and queerness. *Australasian Journal of Philosophy* 62: 111-125.
- Brink, D. 1986. Externalist moral realism. *Southern Journal of Philosophy*, supplementary volume 24: 23-41.
- Carnap, R. 1935. *Philosophy and Logical Syntax*. London: Kegan Paul, Trench, Trubner & Co. Ltd.
- Joyce, R. 1999. Apologizing. *Public Affairs Quarterly* 13: 159-173.
- Kort, L.F. 1975. What is an apology? *Philosophy Research Archives* 1: 80-87.
- Lewis, D. 1969. *Convention*. Oxford: Blackwell.
- Milo, R.D. 1981. Moral indifference. *The Monist* 64: 373-93.

⁸ Something that has been maintained by C.L. Stevenson, R.M. Hare, and P.H. Nowell-Smith, among others.

⁹ I am grateful for the feedback given by David Lewis and Simon Kirchin on an earlier draft on this paper (when it went by the title ‘Noncognitivism, Motivation, and Assertion’). Michael Clark gave extremely helpful comments in the course of revision.

Searle, J.R. 1969. *Speech Acts*. Cambridge: Cambridge University Press.

Stevenson, C.L. 1937. The emotive meaning of ethical terms. *Mind* 46: 14-31.

Stocker, M. 1979. Desiring the bad: an essay in moral psychology. *Journal of Philosophy* 76: 738-753.