Expressivism, Subjectivism and Moral Disagreement

This is the pre-peer-reviewed version of the following article: Expressivism, Subjectivism and Moral Disagreement, Sebastian Köhler, Thought: A Journal of Philosophy Volume/Issue, Copyright © [2012], [Wiley-Blackwell and the Northern Institute of Philosophy], which has been published in final form at http://onlinelibrary.wiley.com/doi/10.1002/tht3.10/full

Abstract

One worry about meta-ethical expressivism is that it reduces to some form of subjectivism. This worry is enforced by subjectivists who argue that subjectivism can explain certain phenomena thought to support expressivism equally well. Recently, authors have started to suggest that subjectivism can take away what has often been seen as expressivism's biggest explanatory advantage, namely expressivism's ability to explain the possibility of moral disagreement. In this paper I will give a response to an argument recently given by Frank Jackson to this conclusion that will show that it is false that subjectivism could explain disagreement as well as expressivism.

Area: Philosophy of Language

0. Introduction

There is a widespread worry about meta-ethical expressivism that it reduces to some form of meta-ethical subjectivism. This worry is enforced by subjectivists who argue that subjectivism can explain certain phenomena equally well that are thought to support expressivism. Recently, authors have started to suggest that subjectivism is actually able to take away what has often been seen as expressivism's biggest explanatory advantage (and a crucial feature that sets expressivism apart from subjectivism), namely expressivism's ability to better explain the possibility of moral disagreement than any form of meta-ethical cognitivism. What such authors have started to suggest is that expressivism and subjectivism are on *a par* when it comes to explaining the possibilities of

disagreement. The thrust of such considerations is well captured in an argument given recently by Frank Jackson (2008):[1] Jackson argues that a standard argument for expressivism, the *Argument from Persistence in Moral Disagreement* (from now on *APM*) supports subjectivism equally well as it does support expressivism and that this shows that subjectivism can explain disagreement as well as expressivism can. In this paper I will give a response to Jackson's argument that will show that the suggestion is false that subjectivism could ever explain disagreement as well as expressivism can. I will first present the APM and Jackson's objection to it and then reply to Jackson's objection.

I. What is the APM?

As mentioned above, the APM is an argument against cognitivism and for expressivism. According to Jackson, *cognitivism* is the theory that moral sentences express beliefs, while *expressivism* is the view that moral sentences express conative attitudes. *Beliefs* are thereby those mental states with – in terms of a familiar metaphor – "mind-to-world" direction of fit, and *conative attitudes* those mental states with – again in terms of the familiar metaphor – "world-to-mind" direction of fit. I will come back to this characterization of cognitivism and expressivism later, but will stick to it for now. I will now present the APM.

The APM has two steps. The first step is to point out that situations are possible in which the parties agree in their beliefs about everything that can be described in non-moral (and non-normative) terms and have made no mistakes in their deductions from these beliefs, but continue to disagree morally. According to Jackson, cases like this are possible and I will follow him in this. The challenge for metaethical theories is to explain how such cases are possible.

This brings us to the second step of the APM: Assume that the facts that are describable in non-moral (and non-normative) terms are all the facts there are. Call this thesis *naturalism* (for lack of a better word). This is a further thesis Jackson accepts, as will I in this paper. Given naturalism, if

1 Jackson's argument is a more developed version of an argument made by Frank Jackson and Philip Pettit in their paper "A Problem for Expressivism" (see Jackson and Pettit 1998: 250/251).

moral facts exist, they must be e.g. in some sense identical to facts describable in non-moral (and non-normative) terms, fully explainable in terms of them, conceptually entailed by them, etc. However, if naturalism is true, only expressivists can explain how cases are possible in which the parties agree on everything describable in non-moral (and non-normative) terms but continue to disagree morally. On an expressivists account this would be possible, because on this account the disagreement between the parties would consist in their having disagreeing conative attitudes, which people can have even if they agree in all their beliefs. Naturalist cognitivists on the other hand couldn't say that such cases are possible, because to say that they disagree morally cognitivists would have to say that it is possible for the parties in such cases to disagree on something describable in non-moral (and non-normative) terms. But, as was stipulated, the parties in such cases agree in *all* their beliefs about what can be described in non-moral (and non-normative) terms. Therefore, on a cognitivist account, they either must agree morally or be talking past each other, in which case their disagreement would only be apparent, not genuine. Hence, cases in which the parties agree on everything describable in non-moral (and non-normative) terms, but continue to disagree morally would be impossible. Therefore, we should prefer expressivism over cognitivism. This is the APM.

II. Jackson's Objection to the APM

According to Jackson, the APM fails even if we grant that cases of persistent moral disagreement are possible and that naturalism is true. It fails because expressivists have failed to specify what they mean by 'disagreement' in the APM. Once we are clear on this we will see that the APM provides no explanatory advantages for expressivism compared to at least one cognitivist view called 'subjectivism'. Jackson's argument is a dilemma, set up in three steps:

The first step is to explain what 'subjectivism' is (see Jackson 2008:77-80). According to Jackson, both expressivism and subjectivism are theories about the relationship between moral

sentences and certain conative attitudes. However, where expressivists say that moral sentences *express* the relevant conative attitudes, *subjectivism* is a cognitivist position that says that moral sentences *report* the relevant conative attitudes, meaning that they express the person's belief that she has the relevant conative attitude. To make the characterization clearer, take someone's utterance “It is impermissible to cheat on one's partner.”. On an expressivist view this utterance would express the speaker's disapproval of cheating on one's partner. On the subjectivists view, however, the speaker would, by uttering this sentence, express her belief that she disapproves of cheating on one's partner.

The second step is to set two conditions which the notion of 'disagreement' expressivists use in the APM must meet for the argument to succeed (see Jackson 2008:80): The *persistence condition,* which says that disagreement in this sense must be able to persist in the face of complete agreement in beliefs about what can be described in non-moral (and non-normative) terms on the assumption that expressivism is true. And the *discrimination condition*, which says that on this understanding of disagreement *only* expressivism and *no* form of cognitivism can say that moral disagreement can persist in the face of complete agreement in beliefs about what can be described in non-moral (and non-normative) terms.

The third step is to ask what expressivists could mean by 'disagreement' when they say that moral disagreement can persist in the face of complete agreement in beliefs about what can be described in non-moral (and non-normative) terms (Jackson 2008:80-82). According to Jackson, two things could be meant by saying that two parties disagree. First, it could mean that they disagree in their beliefs. Two parties disagree in their beliefs if the content of their beliefs can't be true at the same time. Jackson calls this kind of disagreement *factual disagreement*. Secondly, it could mean that they disagree in their conative attitudes. What does this mean? There are different ways to explicate this, but it will not be necessary to go into this issue for the purposes of this paper. It suffices to note that intuitively there are conative attitudes which *can* be in disagreement, for

example intentions with inconsistent contents.[2] Jackson calls disagreement in conative attitudes *attitudinal disagreement*.[3] The conative attitudes which play the relevant role in expressivist and subjectivist accounts will be attitudes that can be in disagreement in this sense.

These three clarifications allow Jackson to set up the following dilemma (Jackson 2008:80-82): When they propose the APM, expressivists could either mean factual or attitudinal disagreement by 'disagreement'. If expressivists meant factual disagreement, then the persistence condition would fail: This is the case, because, according to expressivists, moral sentences don't express beliefs, so that there could never have been factual disagreement in a relevant sense between the parties. Furthermore, in cases relevant for the APM factual disagreement couldn't play the relevant role, since *by definition* the parties agree on everything factual in such cases. If, on the other hand, expressivists meant attitudinal disagreement, then the discrimination condition would fail: On both the expressivist and the subjectivist account the parties will be counted as disagreeing, because, although they agree on everything describable in non-moral (and non-normative) terms, they have disagreeing conative attitudes towards the issue. So, the APM can establish no explanatory advantages of expressivism over subjectivism. Jackson thinks that we shouldn't be surprised that expressivism and subjectivism are on a par with regards to disagreement (Jackson 2008:82). Since the phenomenon of disagreement isn't necessarily tied to language, but is *per se* in the mind, there should be no difference in explanatory success with regards to disagreement between expressivism and subjectivism, since they differ only in their approach to the relationship between mental states and sentences. The question is, what we are to make of Jackson's argument.

III. How Expressivists Should Reply to Jackson's Objection

2 For more elaborate accounts of which conative attitudes can be in disagreement and how, see e.g. Stevenson 1963, Gibbard 2003:48-53, Schroeder 2008:39-44 or Ridge forthcoming.

3 Since Jackson accepts that there can be attitudinal disagreement (Jackson 2008:81), I will not discuss the claim that disagreement in conative attitudes isn't *really* a form of disagreement.

The reply to Jackson's argument I want to propose is to reject his characterization of the explanatory goal of metaethical theories: Expressivism and cognitivism aren't *only* theories about the relation between moral sentences and mental states. Instead they are *primarily* theories about which mental states constitute moral *judgements*. Of course, which mental states are expressed by moral sentences is in the focus of this question, because moral sentences are those sentences normally used to express moral judgements. But, the focus of this approach is broader: It isn't only about what it means to *say* that something is morally right, wrong, etc., but also about what it means to *think* that something is morally right, wrong, etc., to *judge* that something is morally right, wrong, and so on. We should understand expressivism as an explanatory account of both moral language and thought, not only moral language.

Jackson anticipates and discards the way I have proposed to understand the explanatory goal of metaethical theories. According to Jackson, such an approach isn't open to expressivists:

> "For some, myself included, this is an unfortunate way of stating expressivism. For us thinking, judging, and believing are all of a piece in the sense of being representational states, and, in consequence, to think that something *is* good, bad, right or wrong is 'to attribute a property to it' […]. Moreover, it will be obscure how such a view could count as expressivist. To think that X is good, if it is indeed a species of *thinking*, will be a thinking that things are thus and so. But what can the thus and so be in this case other than X's being good? But for expressivists, X's being good is not a way things might be; that's the cognitivists', the opposition's position." (Jackson 2008:79)

What Jackson seems to be arguing here is the following: First, he proposes, by appeal of what he takes to be features of judgements, to understand 'judgements' in terms of belief-like states, that is states with mind-to-world direction of fit. Let us call judgements on this understanding *B-judgements*. Now, an expressivist theory of moral judgements would say that moral judgements are

conative attitudes, that is states with world-to-mind direction of fit. However, it is plausible to assume (and a major motivation for the debate between cognitivists and expressivists) that no mental state can have both directions of fit (see Smith 1994:117-125). If this is true, however, there can be no expressivist account of B-judgements, because B-judgements are *by definition* world-guided states, while an expressivist theory of B-judgements would say that they are world-guiding states. Therefore, there can be no expressivist account of moral B-judgements.

This seems true. But, for Jackson's argument against my proposal to succeed, it must be plausible that when I said that expressivists' focus should be to provide an account of 'moral judgements', I meant moral B-judgements. However, this isn't the case. Instead, I meant the following: There are certain phenomena people call 'moral judgements'. People ask themselves what to do, which actions would be right or wrong, sometimes they come to conclusions on these matters, adopt convictions or change their old ones, they expect others and themselves to behave in certain ways and hold individuals open to blame, resentment and other such feelings, if they fail to do so, etc. Sometimes people express these conclusions by uttering certain sentences and they often use language to guide themselves and others in the process of coming to conclusions on these matters. But, the primary activity seems to be a mental one: reflecting on moral questions and giving answers to them. We could call this mental activity 'moral judgement' – as one would call it in pre-theoretical contexts – or to distinguish it from Jackson's understanding, "moral judgements *qua explananda*" or *moral E-judgements*. It is this mental activity (and its linguistic manifestation) that expressivists want to explain. Now, it is close to a platitude in metaethics that moral E-judgements have a janus-faced nature: They seem to have *both* features of B-judgements and features of conative attitudes. According to my proposal, expressivists would say that the best way to account for this is to assume that moral E-judgements are conative attitudes and then explain why these attitudes behave in ways that allow them the surface features of B-judgements and why it is completely *legitimate* that they behave this way.

Now, for Jackson's argument to go through, he would have to say that moral E-judgement *is* B-judgement. However, he can't *presuppose* this: First, although the phenomenon I have called 'moral E-judgement' is called 'judgement' in pre-theoretical contexts, this doesn't imply that this means the same what certain philosophers mean by 'judgements'. As Lillehammer rightly observes "[c]ommon sense is [...] flexible, with 'judgement' sometimes bearing no cognitive implications beyond the equally neutral 'decision'." (Lillehammer 2002:3). Secondly, although some features of moral E-judgements favour interpreting them as B-judgements (among others, those Jackson himself notes), other features don't. Presupposing that moral E-judgement is B-judgement would therefore beg the question against the expressivist who is trying to take these latter features of moral E-judgements seriously. Of course, Jackson could bring forward an account *why* moral E-judgements are better understood as B-judgements than as conative attitudes. But, this seems to amount to the concession that there is nothing incoherent in an expressivist account of moral E-judgements. And it seems that, *a priori,* nothing speaks against moral E-judgements being conative attitudes. And so, an expressivist account of moral E-judgements is possible. Since this is what I meant by 'moral judgements' when I said that expressivist should aim at giving an explanation of moral judgements, Jackson's argument fails: there is nothing which speaks against holding that the explanatory goal of meta-ethical theories is to provide an account of moral judgements.

Now, on this understanding of the explanatory goal of metaethical theories, *expressivism* would be the theory that making a moral judgement is being in a certain conative state. *Cognitivism* would be the position that making a moral judgement is having a certain belief. And *subjectivism* would be the view that making a moral judgement is having a belief that one has a certain conative attitude. However, if we understand metaethical theories this way, expressivists can escape Jackson's dilemma:

First, they could say that what we are trying to explain is how the parties in the APM could be disagreeing in their moral *judgements*. And this set-up allows expressivists to show why their

theory has an explanatory advantage over cognitivist positions, including subjectivism: What we want to say is that the parties in cases of persistent moral disagreement disagree in their moral *judgements*. Since expressivists say that to make a moral judgement is to be in a certain conative state, that two individuals disagree in their moral judgements should mean on their account that they are disagreeing in conative attitudes. So, on the expressivists view, disagreement in moral judgement *must be* attitudinal disagreement. But, since two individuals can disagree in conative attitudes while in agreement on everything describable in non-moral (and non-normative) terms, it seems that Jackson's persistence condition on this understanding of disagreement is satisfied. On a subjectivist account, on the other hand, to make a moral judgement means to have a belief that one is in a certain conative state. But, on a subjectivist account making a moral judgement *isn't* being in the reported conative state, *but* having the *belief* reporting the conative attitude. So, since moral judgements are *beliefs* on the subjectivist account, disagreement in moral judgement on their account *must be* factual disagreement. However, the beliefs which would be someone's moral judgements on the subjectivist account are her beliefs about her *own* conative attitudes. And, these beliefs can't be in disagreement with the moral judgements of other individuals, because what would make each of them true is something completely different from what would make the others true (namely each individuals own conative attitudes). For example, what makes A's belief that he approves of cheating true are *his* conative attitudes and what makes B's belief that she disapproves of cheating true are *her* conative attitudes. But, this means that the relevant beliefs can always be true together (A's belief that he approves of cheating and B's belief that she disapproves of cheating can obviously be true together), which means that there can be no factual disagreement in the relevant sense on the subjectivist account. If this is correct, then on the notion of disagreement which is disagreement in moral *judgement*, subjectivists can't explain how moral disagreement could persist.

Furthermore, *no* non-subjectivist (naturalist) cognitivist position can make sense of persistent

moral disagreement. Since they would have to say that making a moral judgement is having a belief about facts also describable in non-moral (and non-normative) terms, there should be no genuine disagreements in moral judgements according to these approaches in cases where parties agree on what can be described in non-moral (and non-normative) terms. So, on the understanding of disagreement as disagreement in moral judgement, no form of cognitivism could make sense of the intuition that in the cases relevant for the APM we have a disagreement. Therefore the discrimination condition *is* satisfied. It seems consequently that there is a way for expressivists to escape Jackson's argument, if they reject hid initial set-up.

Let me, in conclusion, note what we learn from my discussion about the ability of subjectivism to steal away expressivism's explanatory advantages when it comes to explaining disagreement. It seems to me that my considerations don't only show that expressivism can better explain the possibility of cases relevant for the APM, but also that the thesis is false that subjectivism and expressivism could ever be on *a par* when it comes to explaining disagremeent. Indeed, it seems to me that in effect, subjectivists are committed to a form of *error theory* about moral disagreement: Subjectivists have to say that although we have the intuition that there is disagreement in moral judgements, the disagreement we encounter is actually disagreement in something else. So, not only can subjectivists not account for the intuition that disagreement in moral judgements could persist, they also have to reject the intuition that moral judgements could ever be in disagreement. On the expressivist account, on the other hand, we can account for these intuitions. So, expressivism is explanatory superior to subjectivism both with regards to the phenomenon of persistence in moral disagreement and with regards to the phenomenon of moral disagreement in general. And this shows that expressivism doesn't reduce to some form of subjectivism, as the claims of the respective theories about the nature of moral *judgement* fall apart.

Bibliography

Gibbard, A. 2003. Thinking how to Live. Cambridge M.A. And London: Harvard University Press

Jackson, F. and Pettit, P. 1998. A Problem for Expressivism. In *Analysis* 58: 239-251

Jackson, F. 2008. The Argument from the Persistence of Moral Disagreement. In *Oxford Studies in Metaethics* 3, ed. Russ Shafer-Landau, 75-86, Oxford: Oxford University Press

Lillehammer, H. 2002. Moral Cognitivism. In *Philosophical Papers* 31: 1-25

Ridge, M. forthcoming. Disagreement. In *Philosophy and Phenomenonological Research*

Schroeder, M. 2008. Being For. Oxford: Oxford University Press

Smith, M. 1994. The Moral Problem. Oxford: Blackwell

Stevenson, C.L. 1963. The Nature of Ethical Disagreement. Reprinted in Foundations of Ethics. An Anthology, ed. Terence Cuneo and Russ Shafer-Landau, 371-375, Oxford: Blackwell