

In Defence of Internalism in the Language of Morals

Adrian Kuźniar

Philosophy of Science Department
Institute of Philosophy, Warsaw University
adrian_kuzniar@uw.edu.pl

Abstract

The principle of internalism states that possessing appropriate motivational attitudes is a necessary condition for the correct linguistic use of moral predicates. It is assumed in the paper that this principle ought to be treated as a standard empirical hypothesis, and that its rational acceptability can be determined only by comparing its explanatory power with that of the rival hypothesis. As it turns out, the principle of internalism is capable of explicating a great many phenomena which continue to confound externalists, while examples of language usage purportedly repudiating it are far less numerous than usually thought.

I

Internalists claim that in order for a particular person's state of mind to be correctly designated as a state of mind consisting in believing that something is morally good or bad, it is necessary that this particular person be in a state of mind consisting in desiring to realize a given act (e.g. [Stevenson, 1937](#); [Gibbard, 1990](#)). Externalists, meanwhile, contradict this stance, claiming that the relation between having certain moral belief and possessing certain desires is at best an empirical regularity (e.g. [Frankena, 1958](#); [Railton, 1986](#)). Thus, say the externalists, while in the real world each moral belief is accompanied by an attendant desire, there exist possible worlds where mental states can rightfully be labelled moral, despite lacking the accompaniment of any desires whatever (cf. [Tresan, 2006](#)).

II

Some of the principal evidence for moral language's possessing the feature of internalism is the sheer surprise, confusion, sense of misunderstanding and desire for an explanation we witness when someone uttering a particular moral thesis fails in opportune circumstances to undertake actions supporting whatever he claimed in his moral enunciations to be good or right, or actions preventing whatever was proclaimed to be bad or wrong. In such situations we take recourse to various hypotheses which explain the observed dissonance: we suspect, for instance, a sudden change of heart, moral conflict, weak will, hypocrisy, wish to provoke, or, as the last resort, communication breakdown, for example addressee's failing hearing or eyesight, or sender's linguistic incompetence. However, if those externalists were in the right who point to wholly incidental nature of the relation between moral judgments and motivations, then such reactions and behaviours of language users would constitute an incomprehensible phenomenon, indeed. Internalists, on the other hand, has no trouble explaining it, for they make possessing appropriate desires one or the sole component of the state of mind which we express in formulating statements involving moral predicates. Matters stand similarly if we consider the regularity noted by Michael Smith (1994, pp. 71-72). Externalists owe us an explanation of why in good and strong-willed individuals a change in moral belief inexorably engenders a change in the subject's motivational sphere. For such a regularity cannot be regarded simply as pure coincidence.

To counter their opponents' objections, externalists can try and lay claim to the concept of a desire to do what is right construed *de dicto*, and not *de re*.¹ Possessing such a primitive, basic, non-derivative and non-instrumental desire would explain why good and strong-willed people react as described above. Hence, whenever such individuals concluded that some act *x* is right, they would discover an inherent motivation to do *x*. This particular motivation would derive from two different "atomic" states of mind: the conviction that *x* is morally right, and the *de dicto* desire to do whatever is morally right (Smith, 1994, 73–6, 1996).

Invoking Bernard Williams's disquisitions on the subject of impartiality, M. Smith 1994, 74–6 propounded a famous argument against the above externalist ruse. It involves accusing externalists of a specific form of fetishism they attribute to people endowed with high moral standing—fetishism that Smith views as wholly inconsistent with who we deem, or ought to deem a morally good person:

Just as it is constitutive of being a good lover that you have direct concern for the person you love, so it is constitutive of being a morally

¹ We have to do with a *de dicto* desire to do what is right when we desire the right actions because they are right, while we possess a *de re* desire when the right actions are desired for their own sake that is, regardless of their moral characteristics.

good person that you have direct concern for what you think is right, where this is read *de re* and not *de dicto*. (Smith, 1994, 76)

III

I find the argumentative force of the argument from fetishism highly questionable. It remains unclear whether the assertion that a good person cannot possess fetishist motivations is Smith's own normative moral judgment, or whether it is a thesis from the domain of descriptive ethics and as such represents people's normative judgments in regard to good moral persons. I assume that it is certainly not any purely conceptual stricture. If so, then it is somewhat difficult to grasp how such arguments invoking fetishism are pertinent to metaethics, to wit a discipline whereby we strive to grasp the meaning and nature of certain expressions belonging to public moral discourse.

Moreover, it appears that at least a sizeable group of competent users of moral language would not consent to excluding anyone from the set of morally good persons solely due to the fact that he desires to do what is right *de dicto*. Let us recall that it is in exactly this type of motivation that Kant (2004, 2005) sought the very essence of morality. Kant's intuitions are shared by Hallvard Lillehammer (1997, 192), when the latter presents the case of a woman who, having at a certain stage in her life grown emotionally indifferent to her own husband, turns up at a party where she begins to experience a *de re* desire to conduct an affair with a handsome gentleman. Ultimately, however, she does not involve herself in any extramarital excursions, precisely because her *de dicto* desire to do what is right prevails, and what she regards as right is remaining faithful to her husband. Does this woman, in the presented excerpt of her moral life, not deserve the label of a morally good person? "No" does not here seem to be the only credible option. Meanwhile, Jonas Olson (2002) believes a morally good person's desire for universal freedom of speech is best explained by invoking the interplay of the person's conviction that freedom of speech is right and his *de dicto* desire to do whatever is right. This is particularly blatant if we were to consider the myriad outrageous opinions to whose dissemination a morally good person has a *de re* aversion.

The subject literature presents us with another line of defence by the externalists. It depends not on neutralizing the accusation of fetishism, but on putting forward alternative models which certainly do not imply fetishism of any kind, and yet permit externalists to explicate reactions typical of good and strong-willed individuals. The best known construction of this sort is James Dreier's (2000) model referring to second order desires.

To sum up, we are forced to admit that externalists dispose of sufficient means of accounting for the link present in good and strong-willed individuals between changes in moral beliefs and alterations in their desires. The situation appears somewhat less rosy if we consider the observable disquiet about moral agents' moral proclamations not corresponding to their conduct. This is because in or-

der to explain the phenomenon externalists would be forced to postulate the presence within the population of a broadly accepted belief in the existence of desires compatible with externalist models. Even if we do share such a belief in regard to the desires of good people, it remains highly dubious whether we hold similar convictions in relation to the mental states characterizing all other participants of moral discourse.

IV

A far more convincing argument undermining the credibility of the externalist claims is the latter's incompatibility with the practical dimension of morality, as it is traditionally interpreted. The key question demarcating the domain of morality is one of Kant's three fundamental queries: *What ought I to do?* (Kant, 2003) Sentences counted as moral utterances always constitute a response to this very question. In formulating the categorical imperative Kant himself leaves no room for doubt as to the imperative, and not descriptive purport of his question and the sought answer. The close connection of sentences containing moral predicates with morality's prescriptive dimension is tersely illustrated by the principle of synderesis: *bonum est faciendum, malum vitandum* ("good must be done, evil avoided"). An uncontroversial formulation of the aspect of morality under discussion would be a principle of prescriptivity according to which inference from sentences containing moral predicates to their corresponding imperatives obtains analytically:

- (1) $\forall x$ [(act x is morally good (right)) \Rightarrow (*prima facie* do x ! / *prima facie* x ought to be done)]
- (2) $\forall x$ [(act x is morally bad (wrong)) \Rightarrow (*prima facie* don't do x ! / *prima facie* x ought not to be done)]

Before we move on, let me stress that states of mind can be incompletely classified into those with a representational function and those with a prescriptive one. The former are to describe the world illustrating how it is in reality, while the latter's task is to influence the world, i.e. indicating how it ought to be. Adopting a technical terminological convention, the first type of mental states shall be labelled beliefs, and the latter—desires. The dispute pitting internalists against their externalist rivals revolves around the role of thus technically construed desires in regard to states of mind which are ordinarily, and not technically, considered moral beliefs. The vexed question is whether such moral beliefs are beliefs in the technical sense.

Let us now ponder what state of mind is expressed by a person using the imperative mood. Clearly, such a person expresses a certain desire she possesses. If, in accordance with the principle of prescriptivity, sentences containing moral predicates analytically entail imperatives, then in every possible world someone's accepting a sentence containing such moral predicates is accompanied by that

someone's having an attendant desire. Yet this is precisely equivalent to accepting the internalist stance! Thus we see that the assumption of morality's practical purport implies the truth of internalist claims, while externalism inevitably leads to what Richard Hare (1981, 71) dubbed "So what?" morality.

V

Arguments against externalism are also provided by the weakness of those metaethical theories compatible with it. It so happens that such theories can only be descriptivist, in other words, must claim that moral language's function is descriptive, and atomic sentences containing moral predicates serve to express states of mind that represent naturalistically or non-naturalistically construed facts.² Besides violating the principle of prescriptivity, descriptivistic conceptions which accept externalism are imperfect in a number of other ways.

Naturalistic versions of externalist descriptivistic theories are incapable of providing an adequate account of moral conflict and of the communication process as it functions in moral discourse. First, moral disagreements frequently persist despite the different parties' accepting identical factual statements and empirical hypotheses. Let the dispute over moral acceptability of capital punishment serve as an example. Let us note that discussion over this issue does not cease even when the different parties agree as to all relevant facts, for instance, that in contemporary social circumstances the death penalty does not deter potential killers, that risk of judicial error is slight, or that modern penitentiary practice makes the likelihood of successful prison break to all intents and purposes negligible.

Second, communication between people adopting divergent, naturalistically determined truth conditions for moral judgments, can continue unperturbed, or in any case it is hardly plausible to compare this situation to instances of purely terminological misunderstanding. A missionary landing on an island inhabited by English speaking cannibals who have nothing else in common with the English culture, shall be able to successfully deploy his language as it stands to speak of good and evil, though the people he labels good or bad would possess completely different character traits than those which inform the cannibals' categorization (Hare, 1952, 148f).

A major flaw of non-naturalistic versions of externalist descriptivistic theories is their inherent incompatibility with the picture of the world as sketched by science. Such theories impute to ordinary language users that in participating in moral discourse, they speak of properties which necessarily fall outside the most adequate picture of the world we have hitherto come up with.

² For the basic division of metaethical theories into descriptivistic and non-descriptivistic, or expressivistic, as they nowadays tend to be termed, cf. Hare 1997, 43–62, Gibbard 2003, 5–8.

VI

Of course, externalists proffer examples of such uses of evaluative expressions which appear to contradict internalist claims. Externalists' most typical and traditionally deployed counterargument is to invoke examples of a motley crew of individuals whose shared characteristic is their morally dubious status. Using Phillipa Foot's (2001, 19f) terminology, one such class of individuals can be termed the shameless. These are persons whose credo is the maxim coined by Ovid in *Metamorphoses*: *video meliora proboque, deteriora sequor* ("I see and approve better things, but follow worse"). Extreme, but real examples of shamelessness thus conceived are people who utter uncontroversial, basic moral judgments of the type: *Killing others is morally wrong*, while being multiple psycho- and sociopathic murderers who show not even the slightest hint of compunction for their crimes. Since one can formulate flawless moral judgments without any volitional proclivity to act on them, so, concludes the externalist, the internalist thesis about moral language must be false.

Internalists, meanwhile, deny that the above-described shameless formulate authentic moral judgments. What they in fact do, is mechanically and thoughtlessly repeat sentences accepted by others or, as Richard Hare put it, use them in inverted commas. In this latter case: "...'I ought to do X' becomes roughly equivalent to 'X is required in order to conform to a standard which people in general accept'." (Hare, 1952, 167)

To what extent is this account of shamelessness detrimental to internalism? I am convinced that, quite to the contrary, it significantly strengthens the internalists' hand. First, the shameless in the strict sense must be distinguished from hypocrites. The former include solely those individuals who in no way try to conceal their motivation to do things which clearly diverge from those they call good and to which they remain wholly indifferent. Let us reserve the term "hypocrite" for all those who in public pronouncing particular moral judgments actually hide their true attitudes and motives. By far the most instances of hypocrisy seem adequately explained by the internalist thesis of inverted commas use of evaluative expressions. Such hypocrites do not speak openly of their attitudes or motives for fear of consequences and sanctions, both formal and informal, by their social surroundings. In order to avoid trouble, they accede to the domain of public morality. It is received wisdom that a like stance is common and well documented. Thus, we can pose an important question: why is the content of moral convictions deemed so significant, immeasurably more so that it engenders the phenomenon of hypocrisy? This is due to the close relation between moral pronouncements and actions, upon which in-group coordination and cooperation on evolutionarily crucial issues is contingent. Moral statements accepted by agents are so relevant to the social setting precisely because they are regularly supported by these agents' motivation to act on them. Thus, we see that explaining hypocritical moral enunciations by invoking the hypothesis of inverted commas use of evaluative expressions actually supports the internalist perspective. In-

deed, inverted commas use appears in the internalists' conception due to the very fact that the language of morals exhibits the feature of internalism—such use is parasitic on this feature of language (cf. Blackburn, 1998, 59–66).

VII

It is of little avail to externalists to resort to a specific class of the shameless: the shameless immoralists. In the main, I shall label as immoralists all those who propagate moral standards contrary to prevailing mores, for instance, earnest and active cannibals in the Western world. These are persons who do not agree with moral judgments commonly accepted in their own cultural setting.

A shameless immoralist is thus someone who feels no compunction about not acting in accordance with his moral judgments, which are contrary to widespread moral beliefs. Such individuals can no longer be said to formulate a statement like, say, *Consuming human flesh is morally right and good!* in inverted commas, as construed by Hare. This is because the judgment is not part of the axiological cultural code accepted within these individuals' civilisational milieu. Is it then not the case that in this and analogous cases, lack of motivation to act on moral judgments is evidence for externalism?

Certainly not in any direct a manner. For even in such circumstances hypocrisy appears a probable explanation for the dissonance, though this time its causes may be different to those outlined earlier. A promising approach seems to be to view the shameless immoralist's proclamations as aiming to shock his social environment, make others regard him as an eccentric, a genius, etc. Signally, a precondition for the elicitation of similar effects is, just as previously, his addressees' universal expectation that moral judgments are followed by actions, an expectation which is a fundamental argument in favour of the internalist stance. Nor is a problem that when resort to the hypocrisy hypothesis, we sometimes invoke human cowardice, and at other times provocative insolence—it is not a problem as people do indeed tend to differ in many respects. But what of cases where the proposed hypocrisy-based hypothesis is obviously spurious in all its variants?

In such instances it is essential to determine how frequently we encounter people who are shameless in the strictest of senses, i.e. shameless individuals who are not hypocrites of any kind. Analysis of this sort is necessary in order to determine the range in which internalists shall be forced to concede that particular moral enunciations are burdened with an error that, unlike with hypocrites, no longer actually supports their position. The larger the group of language users to whom the internalist hypothesis attributes this type of error, the more questionable internalism becomes as a theory which says something about an actual universally and at least implicitly accepted necessary mark of linguistic correctness. However, let us emphasize that the attitude of shamelessness *par excellence* seems to be a very rare and exotic phenomenon indeed. Therefore it does not pose any serious problem for internalists.

VIII

In his article “Internalism and speaker relativism” James Dreier (1990) rejects the above formulated principle of internalism as too strong. Apart from the classic example of a group of sadists who recognize perfectly well what is morally good but are openly motivated to prevent it from happening, Dreier invokes the following case sketched by Michael Stocker:

Recently, I read a story of what might be taken as typical of one course of life. It was said of this political figure that, in his youth, he cared a lot about the suffering of people in all parts of the world and devoted himself to making their lives better. But now he concerns himself only with the lives and fortunes of his close family and friends. He remembers his past, and he knows that there is still a lot he could do to help others. But he no longer has any desire so to do. (Stocker, 1979, 741)

Following Stocker, Dreier thinks the politician could still claim that impartial assistance is morally good without violating rules of moral language. As a result, concluding his investigations on internalism, Dreier (1990, 11) advocates a version of the latter on which the necessary connection between moral beliefs and subjects’ motives is present in *normal cases only*. However, as Dreier himself acknowledges, it is extremely difficult to account for the notion of normality. Neither its statistical rendering (which would undermine internalism as such, according to Dreier), nor grounding it on the contrast with a set of enumerated particular cases believed to diverge from normality (too much of a post hoc procedure to Dreier) seems to be satisfactory; the author of the discussed paper frankly admits that he is not in possession of any adequate analysis of the concept of normality (Dreier, 1990, 14).

In my opinion the quoted example provides no reasons for moderating the internalist stance; nor is it a coincidence that Dreier cannot give a plausible explication of the notion of normality. Let us focus on Stocker’s example. It is easy, I suspect, to notice that the impression that the politician is not transgressing against the language of morals in continuing to label the things he no longer cares about as morally good derives from the context in which he is depicted as helping the needy in his youth. Dreier (1990, 13) himself seems to realize this, conceding that were it not for the context of the politician’s past we would not accept the view that he really knows now that such impartial assistance is morally good. Alas, he incorrectly infers that the example supports a thesis according to which we do have a general principle that allows us systematically to distinguish normal from abnormal cases and thus draw a distinction between cases in which the lack of motivation at least *prima facie* tells against internalism even in its modest version (e.g. a politician who has never possessed a desire to help others but maintains that such help is morally good) and those where it contradicts solely the stronger version of internalism but remains in conformity with its weaker

formulation (e.g. the case of the politician portrayed by Stocker).

This vantage point I take to be confused. That the contexts in question play a vital role in determining whom we judge as competent users of moral language only demonstrates that in our deliberations we pass over important assumptions of the cases entertained. In our initial example of the politician, judging that he could still claim that impartial assistance is morally good, we neglect the assumption stating that the politician currently has no motives whatsoever to help. We do this because the context in which we hark back to his youth strongly suggests that in fact the politician has never ceased to want to help, that that is the sort of man he is; we unwittingly suspect that nowadays he simply lacks sufficient enthusiasm, energy or opportunities to realize his true desires. Thus, we allow for something that is explicitly precluded by the example's premise – and we do so because our life experience teaches us that people's characters tend to remain constant over their lifetime. To sum up: I regard Dreier's argument as resting on an oversight; as such it does not suffice to weaken the internalist position.

IX

Internalism can easily be reconciled with both descriptivistic and non-descriptivistic metaethical theories. Still, its verity seems more supportive of the latter type of metaethical conception. This in turn clears the way for the thesis that what constitutes evidence for an agent's participation in moral discourse is not so much overt, direct invocation of moral values and judgments of what is morally good or bad, but rather his functioning within a particular mode of prescriptivity. Broadly speaking, such understanding of morality implies that engaging in moral discourse is tantamount to accepting and pronouncing norms which *prima facie* demand of given persons to act in a given manner, and of all other agents to effectively impact on those who do not comply. It seems that the mode of prescriptivity in question is deployed on a daily basis in moral as well as legal, religious, aesthetic or narrowly political rhetoric.

References

- Blackburn, S. (1998), *Ruling Passions: A Theory of Practical Reasoning*, Clarendon Press, Oxford.
- Dreier, J. (1990), 'Internalism and speaker relativism', *Ethics* 101(1), 6–26.
- Dreier, J. (2000), 'Dispositions and fetishes: Externalist models of moral motivation', *Philosophy and Phenomenological Research* 61(3), 619–38.
- Foot, P. (2001), *Natural Goodness*, Clarendon Press, Oxford.
- Frankena, W. K. (1958), 'Obligation and motivation in recent moral philosophy', in A.I. Melden, ed., *Essays on Moral Philosophy*, University of Washington Press, Seattle, pp. 40–81.

- Gibbard, A. (1990), *Wise Choices, Apt Feelings: A Theory of Normative Judgment*, Harvard University Press, Cambridge, MA.
- Gibbard, A. (2003), *Thinking How to Live*, Harvard University Press, Cambridge, MA.
- Hare, R. M. (1952), *The Language of Morals*, Clarendon Press, Oxford.
- Hare, R. M. (1981), *Moral Thinking: Its Level, Method, and Point*, Clarendon Press, Oxford.
- Hare, R. M. (1997), *Sorting out Ethics*, Clarendon Press, Oxford.
- Kant, I. (2003), *Critique of Pure Reason*, Palgrave Macmillan, New York.
- Kant, I. (2004), *Critique of Practical Reason*, Dover Publications, Mineola, NY.
- Kant, I. (2005), *Groundwork for the Metaphysics of Morals*, Broadview Press, Peterborough, Ont; Orchard Park, NY.
- Lillehammer, H. (1997), 'Smith on moral fetishism', *Analysis* 57(3), 187–95.
- Olson, J. (2002), 'Are desires *de dicto* fetishistic?', *Inquiry* 45, 89–96.
- Railton, P. (1986), 'Moral realism', *The Philosophical Review* 95(2), 163–207.
- Smith, M. (1994), *The Moral Problem*, Blackwell, Oxford.
- Smith, M. (1996), 'The argument for internalism: Reply to Miller', *Analysis* 56, 175–83.
- 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, 738–53.
- Tresan, J. (2006), '*De dicto* internalist cognitivism', *Noûs* 40(1), 143–65.