

ACCEPTING AGENT CENTRED NORMS: A PROBLEM FOR NON-COGNITIVISTS AND A SUGGESTION FOR SOLVING IT

James Dreier

Non-cognitivist¹ claim to be able to represent normative judgment, and especially moral judgment, as expression of a non-cognitive attitude. There is some reason to worry whether their treatment can incorporate agent centred theories, including much of common sense morality. In this paper I investigate the prospects for a non-cognitivist explanation of what is going on when we subscribe to agent centred theories or norms. The first section frames the issue by focusing on a particularly simple and clear agent centred theory, egoism. The second section poses the difficulty faced by non-cognitivist analyses of such theories and norms, and runs briefly through a couple of abortive attempts to solve it. The third section offers a solution and explains it. The fourth section uses the account developed in the third section to show in what way agent centred judgments are universalizable and in what way they are not.

I. Egoism and Non-cognitivism

It is sometimes thought that non-cognitivist meta-ethics has no normative implications, that it is compatible with any understandable normative view. Here is an argument² to show that non-cognitivism entails the incoherence of accepting egoism.³ Suppose that Tom, Dick, and Harry are playing a game of chinese checkers, with the winner to get a substantial prize. An egoist says: each man should further his own interests, without regard to the interests of others. The egoist is then committed to each of the following:

I came up with the main ideas for this paper as a result of some conversations with Michael Smith while I was visiting at Monash University. I thank him, and also the participants in a faculty discussion group at Brown, especially Dan Brock, David Estlund, and Ernest Sosa.

- ¹ I will take non-cognitivism to be the class of theories according to which moral judgments express some non-cognitive attitude, rather than expressing a belief.
- ² Due to Brian Medlin [12]. As Medlin deploys the argument, it is an argument against the coherence of egoism, taking non-cognitivism as a premise. In 1957, this was a rhetorically reasonable strategy. The more neutral way to think of it is as an attempt to demonstrate the incompatibility of egoism with non-cognitivism. Nowadays, the upshot might well be taken to be a difficulty for non-cognitivist meta-ethics. In the text I extend the argument to point to a tension between non-cognitivism and agent centred normative theories in general. If agent centred normative theories are incompatible with non-cognitivism, I think most moral philosophers will count this incompatibility a deficit of non-cognitivism.
- ³ By 'egoism', I mean the view that says that each person ought to do what is in her own self-interest; I sometimes assume, more for ease of exposition than for any substantive reason, that a person's self-interest consists just in her happiness, so that egoism instructs each of us to maximize our own happiness. By contrast, I do *not* call 'egoism' the view held by a person that says that everyone ought to act to fulfill that person's self-interest. I will suggest another name for this view. Nor am I considering at all another view sometimes called 'egoism', namely, that we each do in fact always act to maximize the satisfaction of our interests.

- (T) Tom should win the game.
- (D) Dick should win the game.
- (H) Harry should win the game.

These judgments, according to the non-cognitivist, express non-cognitive attitudes, which for now we may just call 'desires'. Which desires do they express? These:

- (t) The desire that Tom win.
- (d) The desire that Dick win.
- (h) The desire that Harry win.

But if the egoist is expressing those three desires, then he is expressing incoherent desires, for he knows it is not possible that more than one player win.⁴

I take this to be the core of the argument given in Medlin [12]. My plan in the next section, is to sharpen the argument a bit, providing more motivation for attributing certain interpretations of normative judgment to non-cognitivists and then to extend the incompatibility thesis from egoism proper to the general category of agent centred normative views. I'll argue that the incompatibility thesis is a problem for non-cognitivism.

II. The Incompatibility Thesis

Here is the thesis:

- (IT) If non-cognitivism gives the correct analysis of moral judgment, then egoism (and any agent centred normative view) is incoherent.

I will present an argument for this thesis, and I will claim that if true it poses a serious problem for non-cognitivism, for egoism, and especially other agent centred normative views, are not, in fact, incoherent. Ultimately, I will be trying to answer this argument on behalf of non-cognitivism.

In presenting the argument, I am going to assume that judgments of what someone ought to do are backed, or mirrored, by judgments of what is good. One ought to do just what has the best results. This assumption is the assumption of teleology. It is a controversial assumption, even in the restricted context of egoistic theories. Henry Sidgwick, famously, believed that an egoist could be driven from his theoretical position if he made claims about what is good, but not if he restricted his claims to judgments of what people ought to do. Nonetheless, I think the assumption of teleology is innocuous here. And it is helpful: assuming teleology helps me to present the problem of compatibility in a stark way, and will also help when I offer my solution.

⁴ Of course, this is questionable. It is questionable whether there is anything incoherent in having a bunch of desires which you know not to be jointly satisfiable. But I will eliminate this loophole below, by replacing talk of desires with talk of preferences; at the moment I am just setting the stage.

According to a non-cognitivist theory, what attitude, what state of mind, is expressed by a moral judgment in general?⁵ One complication is that some non-cognitivists, notably R. M. Hare, say that moral judgments express prescriptions, not states of mind. But our ultimate question arises for Hare's version, too, for he will be obliged to say what it is that constitutes sincere acceptance of a moral judgment or principle or theory. So let's call this the problem of explaining the 'Acceptance State' (AS) of a given moral judgment, principle, or theory. I believe we should follow closely what Hare does say about the AS.

As sincerely accepting a statement involves believing it, Hare says, sincerely accepting a prescription involves acting on it 'on the appropriate occasion and if it is within our power.'⁶ Later, Hare explains that prescribing something entails *preferring* it,⁷ and says that 'to have a preference is to accept a prescription.'⁸ The reason for associating moral judgment (which Hare thinks of as a kind of prescription) with preference, rather than with, say, some affective state, is simple. The motivation for a non-cognitivist theory of normative judgment comes from the special connection between normative judgment and action. Normative reasoning is practical reasoning; when it functions ideally, it issues in action. So the sort of state it expresses, the sort of state that constitutes its sincere acceptance, has to be a state that of its nature motivates.⁹

Let's return to egoism. The egoist judges that Tom should advance his own interests even at the expense of Dick's, and also that Dick should advance his own interests even at the expense of Tom's. We might now decide that the ASs for these two judgments are,

(AS₁) Preference for Dick's interests over Tom's

and

(AS₂) Preference for Tom's interests over Dick's

⁵ Notice that the problem arises not only for proper non-cognitivists, but equally for any theory with a non-cognitivist 'core.' A theory has a non-cognitivist core if it explains the nature, content, or the like, of moral judgment by reference to the non-cognitive attitudes of someone who accepts moral norms. For example, an Ideal Observer theory, according to which moral judgments have factual content about how a suitably characterized observer would judge the situation at hand, will need to say what that observer is expressing. Clearly, it won't do to analyze a judgment that something is morally good into the judgment that an Ideal Observer would believe that it is good. The analysis must be in terms of some other attitude of the Ideal Observer than belief, and typically proponents have suggested a non-cognitive attitude; for example, Firth [3] mentions 'deciding in favor' of one act. And similar remarks apply to Moral Sense theories, and to certain kinds of relativism.

⁶ Hare [5] p. 20.

⁷ Hare [6] pp. 21-2.

⁸ p. 91.

⁹ For an elaboration of this 'internalist' thesis, see my Dreier [2]. There I also explain what it is about certain preferences that makes them moral preferences, what it is about moral judgments that distinguishes them from other sorts of normative judgments. I favour what Allan Gibbard calls 'cluster constraints', according to which a number of separable factors contribute to the moral quality of a judgment. Judgments that meet some constraints but not others may be 'somewhat moral', and we should just admit that it may be vague whether a given judgment is moral or not.

And surely these two states are inconsistent, in whatever sense preferences can be inconsistent. But is there any reason to think that these are the ASs for the egoist's judgments? Yes, there is some reason. Recall that we are assuming teleology. Reasons for performing one act rather than an alternative come from the first act's having better consequences than the second. What should the AS be for a judgment of the form, 'A is better than B'? If ASs are preferences, then the natural account would be that that judgment has an acceptance state of preferring A to B. In fact, since the preference relation of decision theory is logically very much like the 'better than' relation, we can give a natural rule for constructing an AS for any given (teleological) moral theory.

(Rule) Take the 'better than' relation for the theory, and construct the preference ordering with just that relation as its preference relation. The preference ordering will be the AS for the theory.¹⁰

But if that is the general account of ASs, then the incompatibility between non-cognitivism and egoism should be apparent. To simplify, let's hereafter consider a game of two players. Suppose Alan and Josie are playing a game of chess with a substantial prize for the winner. The egoist onlooker says that Josie ought to bring it about that she wins, and that Alan ought to bring it about that he wins.¹¹ So he seems to be committed both to saying that it is better that Alan win than that Josie win, and also better that Josie win than that Alan win. Now, there is already an apparent contradiction, but of course the point the egoist will make is that the 'better than's in the two judgments are not the same relation. When egoism ranks states of affairs, it produces not a single ordering but an ordering for each agent. Goodness, according to egoism, is agent relative, or as I will say, agent centred.¹² But precisely because it is agent centred, egoism cannot be assigned an AS according to our Rule. For the Rule tells us to take *the* 'better than' relation for the theory, and there is no unique one for egoism. And that is the crux of the difficulty.

Note that the incompatibility between non-cognitivism (coupled with the Rule for assigning ASs to moral theories) and egoism arises because of an easy-to-see structural feature of egoism, its agent centredness. Before investigating possible solutions to the problem, let's look at what this amounts to. Moral theories tell people what to do, and teleological ones tell us especially what to bring about. In accordance with universalizability, their advice must be put in a general way, addressing 'the same' advice to each agent. That is to say, the advice will have the form,

(ADV) (x) (x is to bring it about that)

⁹ *Continued*

Gibbard rejects this view, which he raises as an alternative to his own. See Gibbard [4], esp. pp. 203 f. In the present paper, I merely assume for simplicity that moral preferences are those one has from the 'moral point of view', and that any plausible non-cognitivism will have to say what characterizes that point of view, since it is an intuitively recognizable idea.

¹⁰ The match is a good one. A teleological moral theory is characterized by its 'better than' relation, in the sense that two theories with the same 'better than' relation will be difficult to distinguish, and arguably mere notational variants of each other.

¹¹ Of course, these two 'oughts' cannot be realized together, but there is no cause for alarm merely in *that* incompatibility. See Humberstone [7] and the works cited there in the first note.

¹² Scheffler [16] uses the expression 'agent centered,' which I prefer because no one will confuse it with other sorts of 'relativity.'

where ‘ x ’ is replaced, in a particular theory, by a formula containing at most one free variable, ‘ x ’. If the theory is purely agent neutral, then each bit of advice will contain an x -formula with no free variables (nor any anaphoric reference back to x , as in ‘her interests are advanced’). If some of the advice contains free occurrences of ‘ x ’, then that advice is agent centred, and (so as to have only two categories) we call the theory ‘agent centred’.¹³ When a theory is agent centred, so is its ranking of outcomes. If every x is to bring it about that x ’s interests are advanced, then the ranking of outcomes according to how well they satisfy interests will be different depending on who is doing the ranking. And similarly for other agent centred advice and theories. The point is that the incompatibility between egoism and non-cognitivism will also be an incompatibility between any agent centred theory and non-cognitivism. That is why I said that any incompatibility is better thought of as a problem for non-cognitivism than as a problem for egoism. Nowadays, at least, non-cognitivists tend to be quietist about ordinary moral thought.¹⁴ It might not be terribly revisionary of ordinary moral thinking to toss egoism out of the realm of intelligible moral views, but agent centred elements in general won’t go quietly.¹⁵

Now to canvass a couple of quick attempts to find a replacement AS for agent centred moral views. Since there are too many ‘better than’ relations at hand to apply our Rule, we might try either (a) Incorporation—generating a preference relation that includes all of the various ‘better than’ relations assigned to agents by a theory; or (b) Selection—choosing some privileged one for the AS preference schedule to match. It is easy to see that Incorporation is doomed. The many rankings assigned by an agent centred theory will in general conflict with each other, not in every pairwise comparison, but often enough. As we saw, egoism assigns to Alan a ‘better than’ relation that ranks his winning above Josie’s winning, and to Josie a relation that reverses that order. These two relations cannot be combined into a single grand one, on pain of incoherence. So Incorporation is out of the question.

Selection may look more promising. We might seize the following fact to our advantage: for practical purposes, an egoist needn’t worry about how outcomes and acts are ranked from the perspective of others. So maybe the AS for egoism is the preference schedule that matches the ordering imposed by what is better or worse from his perspective, the egoist’s own. Let’s look at an example.

Since Alan is an egoist, he prefers winning the chess game to losing it. And, if asked, he would certainly say that the state of affairs in which he wins is better than the state of affairs in which Josie wins. In virtue of what, according to a non-cognitivist, is Alan an egoist? Perhaps just in this: that he prefers A to B just in case egoism says that A is better than B from Alan’s perspective. Certainly, if we could peek at Alan’s preference relation and discover that it matches exactly the relation ‘is better for Alan than’, and we were asked what sort of ethical theory Alan accepts in virtue of these preferences, the tempting answer would be that Alan accepts

¹³ Nagel [13] gives a formal distinction between objective and subjective reasons that is similar to my distinction between agent neutral and agent centred advice and theories, and Kagan [8] and Scheffler [16] use distinctions in the same spirit. See also Sen [18].

¹⁴ Most famously Blackburn [1], Chapter Six, the *locus classicus quietus*.

¹⁵ See, for example, Nagel [14], Chapter Nine, for agent centred pieces of common sense morality. There he says that agent neutral reasons in ethics are about what should happen, and agent centred ones can be about what people should do (p. 165). But this seems wrong. See Sen [17] pp. 29-31, and also Sen [18] p. 120 and note 15 therein.

egoism. Tempting, but, I think, not adequate. The difficulty is that if Alan's preference schedule is taken to be the AS for egoism, then we won't be able to distinguish egoism from another theory, one that is related but distinct.

Contrast egoism with egomania. An egoist thinks that he ought to be maximizing his advantage, and that you ought to be maximizing yours. When Josie says that it will be better if she wins, Alan will think that she has judged correctly, even though he will say it's better if he wins. But suppose Alan were an egomaniac instead. Then he would believe that everyone should be seeking Alan's advantage. 'If only they understood right morals,' he would muse, 'all those fools would stop being so selfish and devote themselves full time to me.' Even an egoist would be shocked! Intuitively, two egoists share a moral theory, but two egomaniacs have different moral theories. Our Selection proposal conflates egoism with egomania. Given the preference schedule we tentatively proposed for egoism's AS, we cannot tell whether Alan is an egoist or an Alanist, an egomaniac.

It might be thought that this conflation is one a quietist non-cognitivist can live with. The distinction between egoism and egomania is relatively arcane. Maybe it is the product of excessive theorizing, and not a rooted shrub in the landscape of common sense moralizing. But, again, there will be parallel confluences for our Rule's assignment of ASs to other agent centred theories. For example, it is commonly thought worse for a person to break a promise than for that person to do something that causes another person to break her promise. Suppose John and Liz have each promised to send me a postcard. John knows that if he sends me a postcard, Liz will not bother to send me one, but if he doesn't, then she will send me one. It is worse for him to break his promise, thereby causing Liz to keep hers, than it is for him to keep his, thereby causing Liz to break hers. Or anyway, suppose this is so. The promise-keeping principle, then, offers this advice:

(PK) (x) (x is to bring it about that x's promises are kept.)

The AS for the promise-keeping principle is then a preference that one's own promises be kept (that is, preferring their being kept to their being broken). If John accepts the promise-keeping principle, according to our Rule, that means he prefers his promises to be kept. But now consider an odd principle giving this advice:

(PK_J) (x) (x is to bring it about that John's promises are kept.)

This principle gives a special importance to John's promises from an agent neutral perspective. Can't John accept the commonplace promise-keeping principle without accepting (PK_J)? It would be unfortunate, I think, if non-cognitivism were to entail that there is no difference between subscribing to the ordinary promise-keeping principle and subscribing to a principle directing everyone to give special weight to your promises. Unfortunate, that is, for non-cognitivism. So Selection is not a satisfactory solution.

It is worth considering one more defense, what I'll call the 'Trying Defence.' Someone might insist that telling agents to bring about the good is poor advice. Better to tell each agent to *try* to bring about the good. For it may not be in an agent's power to succeed in bringing it about that she is happy, or that a famine ends,

but it is always in an agent's power to try. A normative theory has to tell an agent to do things that are in that agent's power, otherwise it isn't giving *advice* strictly speaking. I have some sympathy with this point.

Suppose egoism told each person to *try* to bring about her own happiness. Could the simple Rule then be applied coherently? On the face of it, the difficulties besetting the egoism that assigns different and incompatible better than relations to each agent are overcome by this Trying strategy. For there is nothing incoherent about my wanting you to try to beat me in chess, even though I do not want you to succeed. I want to succeed, and I want to try, but I want you to try, too. No incoherence there.

Jesse Kalin notes the existence of the Trying Defence,¹⁶ but says that the Trying form of egoism is just a different theory. And so it is. To see this, consider the theory that tells each of us that what is of primary value is his own welfare, and then a lexically lower, agent neutral value is in the effort of each person to pursue his own welfare. Such a theory is a hybrid, with an agent centred component and an agent neutral one. It is distinct from egoism proper, which has only an agent centred component.

The hybrid theory is somewhat peculiar. Notice that since my own welfare has a lexically higher value, relative to me, compared to the value of the tryings of others, I will not in general have occasion to act on the agent neutral valuation of the general effort of each to pursue his own welfare. Would I ever have occasion to act on it? I might, if I were in a situation in which I could bring it about that someone tries to pursue her own welfare, and none of my own welfare is at stake. But it seems to me that these occasions would be few. Pursuit of one's own welfare is a fairly demanding project, leaving little time for other, lesser values. So the hybrid theory adds to the pure version a value that we would rarely have occasion to pursue. Never mind; I insist only that the hybrid theory is not the same as the pure version. So the Trying Defence is not, in the end, a defence of the ability of the Rule to make egoism safe for non-cognitivism.

III. A Solution

A hint at the solution comes from some ordinary ways of describing preferences. Remember that the problem arises when a theory gives not a single betterness relation, but one for each agent, and that each agent appropriately accepts the betterness relation ascribed to him by the theory when he takes that relation as his preference relation. But there is no way to ascribe to each egoist *the same* preferences. Or is there? Certainly there is. Josie and Alan each want something different: he wants it to be the case that Alan wins, and she wants it to be the case the Josie wins. Then again, they each want the same thing: Alan wants to win, and so does Josie. Speaking in ordinary untechnical English, it is very easy to enunciate a sense in which all egoists have just the same preferences. They all prefer more happiness to less. Put a bit more technically, each prefers alternative *A* to alternative *B* just in case he is happier in alternative *A* than he is in alternative *B*. Do they have the same preferences, or not? That depends on what the objects of preference are taken to be.

Preference is, as a general rule, harmlessly thought of as a propositional attitude. The phenomenon we've just noticed in the domain of preference is familiar in the domains of other propositional object verbs. Professor Speck tells her colleague Blob: 'I have just informed the Dean that I will happily serve on the committee.' Blob

¹⁶ See fn. 10 of Kalin [9].

replies, ‘Oh, so have I.’ Has Blob told the Dean that Blob will happily serve, or that Speck will happily serve? What Blob said is ambiguous. I expect to be the first one sacked in the coming recession, and so do you. Or, by contrast, I expect that I will be the first one sacked, and so do you. In each case, we expect the same thing, only the kind of thing (‘to be the first one sacked,’ ‘that I will be the first one sacked’) is different in each case.

When what you desire, or prefer, or expect, or ..., is something of the form *that p*, then we can speak of the object as a proposition. More helpfully, keep in mind the linguistic form that expresses them, namely, a sentence. When what you expect or prefer is something of the form *to Φ*, then what is the object? What is expressed by the infinitive? With no general theory of what is expressed by a verb in infinitive form, we can even so think of the object as a proposition minus an agent. (I want to eat; I want Jordan to eat.) When Linda wants to *Φ*, the object is the proposition that Linda *Φ*s, minus Linda herself. Doesn’t she want the whole proposition? Why leave *her* out of it? Well, we want it to turn out that when Linda wants to *Φ*, and Leslie wants to *Φ*, they want the same thing. The thing that they both want is the proposition with the agent removed.

There is a ready assimilation of talk of property desires to the well-developed apparatus of indexicals and their role in ascriptions of propositional attitudes. We might think of Perry’s “essentially indexical belief,”¹⁷ or Kaplan’s characters.¹⁸ But things will be clearest, I think, if we adopt David Lewis’ scheme from ‘Attitudes *De Dicto* and *De Se*.’¹⁹ In the case of believing, there is believing a proposition, and also believing a property. Believing a property is what Lewis calls ‘self-ascribing’ the property. Similarly, preferring a property (preferring, that is, to instantiate it) we can call ‘self-prescribing’ it. (A doctor might self-prescribe ‘taking two tablets daily,’ and that’s just the sort of thing we mean.) Talk of properties here is meant in a loose sense; there is no thought of real universals or privileged classes or natural joints. Any old open formula (with one variable free — the one that the agent will instantiate if the self-prescription is satisfied) expresses a property.

Let us be explicit now. Agent centred theories can’t be summarized by a single betterness ordering for states of affairs, but they can be summarized by a single betterness ordering for properties. In many cases the betterness ordering in question will be obvious. Take egoism again. If we have a scale of happiness available, we can put the betterness ordering this way: being happy to degree *j* is better than being happy to degree *k*, just in case $j > k$. Being happy to a certain degree is a property. It is expressed by this open formula:

(OF) x is happy to degree j .²⁰

Accepting an agent centred theory is taking its betterness ordering as one’s preference ordering, where preferences are construed as self-prescriptions.

¹⁷ Perry [15].

¹⁸ Kaplan [10].

¹⁹ Lewis [11].

²⁰ Of course, ‘*j*’ is here thought of as a constant, and ‘*x*’ is the free variable. Some may prefer:

x (x is happy to degree j)

as a name for the property, or a predicate.

I claim that the non-cognitivist notion of accepting a normative theory is properly cashed out in this way, whether the theory is agent centred or not. And this claim can be factored into two: first, that orderings of properties individuate normative theories successfully, so that no two distinct theories are mapped to the same ordering; and second that when a person's preference relation (over properties) is the betterness relation (for properties) of a normative theory, it is intuitively right to say that the person accepts the theory. I will argue for the two factors below.

Talk of self-prescribing evokes R. M. Hare's Prescriptivism again. And a good thing, too, for it will help to borrow one of Hare's devices. Hare takes moral judgments to be a kind of prescription, and he takes it to be a logically essential feature of moral judgments that they are universalizable. By this he means that when making a moral judgment, I must be willing to prescribe just the same for other cases than the actual one which are just like the actual one in universal respects (meaning, in respects other than those of numerical identity of the individuals involved). So, when I say what's best here and now and actually, I have to be willing to say that the same thing would also be best if I were the one playing the black pieces, instead of the one playing the white pieces. And that means I must prescribe that the player of black should win in both the actual and the merely possible situation, or that the player of white should win in both, or that the game should be a draw in both (or, I suppose, I could simply be indifferent about the outcome – but this is scarcely believable!). It is worth noticing that an egoist cannot meet this condition. Alan prescribes that white should win in the actual situation (since he's the one playing white), but for the possible situation just like the actual one except that he is the one who is playing black (and has every property actually had by the one actually playing black, that is, all of Josie's properties) he prescribes that black should win. And of course his preferences shift around in the parallel way. So egoism is not a universalizable preference, so it doesn't count as a moral theory, according to Hare. We shall later on give an alternative characterization of universalizability on which egoism passes. But for now let's see how our new kinds of preferences, the Lewis-inspired preferences *de se* figure in Hare's sort of examples.

The worry is that egoism will turn out to be the same as egomania in its AS. For Alan prefers that white win, just in case Alan is playing white, and that black win, just in case Alan is playing black. He does have a standing preference that he expresses thus: 'I prefer that I win.' And he also has this standing preference: that Alan win. No accident, for he believes that he is Alan. So, for every situation, whether actual or possible, whatever he wishes for himself he wishes for Alan, and vice versa. So it is hard to see how there are going to be any egomaniacal preferences of his that aren't egoistic preferences, nor any egoistic ones that aren't egomaniacal. So it does not appear that any preferences will be around to determine whether he is an egomaniac or an egoist.

But preferences *de se* rescue the two theories from conflation. If he is an egoist, Alan will self-prescribe this property: winning. If he is an egomaniac, Alan will self-prescribe this very different property: being such that Alan wins. It should be manifest that these properties are different. Suppose Alan loses to Josie. Then Josie has the first property and lacks the second. It is only Alan himself who manages to have the first property if and only if he has the second. They are different properties, all right, but one might still worry that the preferences

have not been distinguished in the proper way, since after all, Alan has the first property, the one that is self-prescribed by all the egoists, if and only if he has the second, the one that is self-prescribed by all the Alan-enthusiasts (and especially by Alan himself if he is an egomaniac). Properties that are necessarily instantiated or uninstantiated together by Alan might be undifferentiable *as targets of self-prescription by Alan*. They are different properties, but the question is, do self-ascriptions of them by Alan really amount to different preferences?

In the actual world, egomania and egoism look very much alike. That's because the two are coextensive in the presence of a certain belief. For Alan, it is the belief expressed, 'I am Alan.' It is the self-ascription of being Alan. And, importantly, he believes it with what amounts to certainty. In the presence of that belief, virtually every disposition that the one preference participates in, the other does too, but that hardly means that the preferences are the same preference! If Alan is an egoist, he doesn't prove by his actions that he is not an egomaniac. But for that matter, neither does he prove by his actions that he is not a Bonapartist (maybe he believes he is Napoleon). And you won't learn what normative theory he accepts by asking him about what states of affairs are good, either. But you could find out by asking him about what is right, or by asking him about which properties are good.

A realist (who is also a teleologist) might put it this way. What makes an action right is that it is in pursuit of the good. We should take attitudes *de se* as primary, constructing the *de dicto* sort out of them as the need arises. So the good in the pursuit of which all right action consists, one might say, is some property. Alan the egoist believes that the good is happiness (that is, the property: being happy, not, the state of affairs: that people are happy). Alan the utilitarian believes that the good is being in a happy neighborhood (or world).²¹ Alan the egomaniac believes that the good is being such that Alan is happy. I think this way of distinguishing the three views is very intuitive. And notice that the distinguishing beliefs carry over easily into non-cognitivist descriptions, for although non-cognitivists do not say that the distinguishing attitudes are *strictu dictu* beliefs, the conative counterparts are readily available. Believing something good, according to the non-cognitivist, is properly understood as preferring it.

Now for the second part of my claim: that identifying acceptance of a normative theory with having the theory's betterness relation over properties as one's preference relation over properties squares with our intuitive thinking about what is involved in accepting a normative theory. Accepting egoism turns out to be having the

²¹ Ernest Sosa pointed out a potential difficulty with regarding a utilitarian preference as a preference for a property. The problem is that a preference for a property is a preference to instantiate the property, and for you to instantiate a property you must exist. But a utilitarian's preference does not typically involve her own existence. Put in the terms used in the part of text to which this is a note, the problem is that a utilitarian does not so much care about *living* in a happy world, as he cares that *this* world be a happy one.

I have to admit that I find this issue extremely confusing. For example, it seems to me that in order to decide whether this objection makes sense we have to decide whether modal realism and counterpart theory are correct. And that task is beyond me, and certainly beyond the scope of this paper. Fortunately, even if the problem is a genuine one, it is not serious for my purposes. It matters to the question of whether preferences *de dicto* are reducible to preferences *de se* (as Lewis [11] says they are), but if they aren't it matters little to the main point here. Perhaps agent neutral theories have ASs that are preferences irreducibly *de dicto* and agent centered theories have ASs that are irreducibly *de se*. Then some theoretic simplicity is lost, but the main thesis can stand the loss.

preference to be happy; accepting utilitarianism turns out to be preferring to live in a world in which there is more happiness. When the utilitarian says that a certain action is right, does he express his preference for that action? The utilitarian always does prefer the right action, he always does prefer the action that produces the most happiness of those acts that are available. But the egoist does not always prefer the right action, that is, he does not always prefer that a given person perform the right action on that occasion. In fact, he rarely prefers it, since when someone else does the right thing, that rarely turns out to maximize the egoist's happiness. So, what *does* the egoist mean by calling it right? When Alan says that Josie made the right move, he isn't going to be happy about her move. He wishes she'd made the wrong move, but he thinks she did right in making the move she did, because (let's say) with that move she guaranteed victory for herself. The right act is the one by which the actor comes to bear a good property (the property of being the winner, for example). When Alan says Josie's act was right, he is approving of something; he is approving of the target of the act, a property. Calling an act right is expressing approval of (preference for) its target. So the sets of preferences I am assigning as the ASs of moral judgments and theories do square with our intuitions about what people who accept those judgments and theories will prefer and approve.

Non-cognitivism seeks to explain how a normative judgment can, and must, by its nature, express an attitude of approval. How is it that calling something right is essentially a way of recommending it? According to the present proposal, when I say that it is right to keep one's promises, I am expressing my favourable attitude toward the target of that sort of act. The difficulty is to make out the difference between the attitudes expressed by (PK) and (PK_J). But we are now in a good position to meet this difficulty. The AS for (PK) is the preference for being a promise-keeper. The AS for (PK_J) is the preference for being such that John's promises are kept (or more naturally put, the preference that John's promises be kept). Resort to property preferences enables us to capture the idea that resides at the center of non-cognitivism, even when the normative judgments we are modeling are agent centred ones.

IV. Universalizability

We're now in a position to understand in what sense an egoist's judgments are universalizable, and in what sense not.

It is said that to be making a *moral* judgment, I must be willing to universalize my judgment. What does this mean? As a first try, we might think of a universal judgment as one containing no indexicals or proper names. Your judgment is a candidate for a moral judgment only if you are willing to accept some paraphrase in which every name and indexical has been exchanged for some description. It is plausible to put indexicals and proper names together in a basket, because they share a relevant feature, namely, that they are directly referential. Names and indexicals contribute a particular to propositions expressed by sentences in which they occur. The constraint of universalizability is designed to capture the intuitive thought that moral judgments are not supposed to be judgments essentially wedded to any particular, especially not to any particular person. This is the sort of constraint Hare is shooting for in Hare [6], as is shown by the thought experiment he offers as a test of universalizability. I say that it is wrong for my money to be redistributed to the less well off. Hare asks

whether I can sincerely and with full awareness of relevant information also say that the redistribution would be wrong in the merely possible situation just like the actual one except that it's me who is the needy one, and someone else the better off. If I cannot, my judgment has failed the universalizability test. If the original moral judgment was *essentially* indexical, if the reason I judged the redistribution wrong was that it was *me* who was losing out, then that was no moral judgment at all.

According to the present account of acceptance of norms, there is something especially important about judgments that are ineliminably indexical. For how should we understand a preference in favor of an essentially indexical proposition?²² We might think of it as a preference toward a *character*, to use Kaplan's term.²³ A character is a function from contexts to contents, and the kinds of sentences relevant here take the agent in the context (the speaker, in Kaplan's index) as the argument for their characters. But these functions are nothing more than properties of just the sort we are using in our theory. When we have preferences *de se*, when we self-prescribe properties, what our preference relation ranks are open formulas. And we could just as well replace these with indexical sentences; thus, the property of being happy, expressed by 'x is happy,' gets replaced with 'I am happy,' expressing the proposition that I am happy.

There is something amiss in the formulation of Hare's test. When we ask someone what he prefers regarding the possible situation *exactly* like the actual one except that roles are swapped, what are we asking him? Is there really a different situation like that? If there is a possible situation containing someone with all of my properties, isn't that someone *me*, and if the situation is like the actual one in all universal respects, isn't it the actual one? Is there, in fact, any way for situations to differ above and beyond their universal respects? We needn't think so. When a situation is described and I am asked what I prefer regarding it, I may properly ask, 'Which one is me?' And I am not asking for any further specification of the situation. I may wonder which is me even when I know exactly which world is described.²⁴ And if I am an egoist, that is (nearly) always a relevant thing to wonder.

So it will come as no surprise that egoism doesn't pass the universalization test. When Alan considers the possible situation just like the actual one, except that he occupies Josie's actual role (and she his), he does not accept the prescription he accepts regarding the actual situation, because he does not prefer the same universal proposition for that situation. Actually, *regarding* the actual situation, he prefers that white win. Actually, regarding the possible situation, he prefers that black win. Of course, for each situation he prefers that *he* win, but that is to say he prefers a property, not a proposition. Which property preferences can pass universalizability? Those whose objects are properties necessarily shared by worldmates. For example, the property of living in a happy world is one which is shared by all worldmates of anyone who has it. Preferences

²² Here I am pretending that there are propositions, and not just sentences, that are indexical. Maybe this is not a pretence; maybe propositions really can be indexical. Maybe, but I don't believe it. In any case, it will become clear that the loose way of speaking here is harmless, because in the end we think of the preference expressed by an indexical sentence in a way that presupposes nothing in particular about what sorts of things can be indexical.

²³ Kaplan [10].

²⁴ As Lewis shows in the example of the Two Gods, Lewis [11].

whose objects are properties that some may have even though others lack them are not universalizable.²⁵

Let us check the results against the promise-keeping principle. The commonsense (PK) cannot pass Hare's test of universalizability. In the actual circumstances, John prescribes that he, the husband, send me a postcard, though he knows that this will cause Liz not to send me one. For John's obligation is to keep his promise, not to maximize the number of promises kept. But let him consider the situation just like the actual one, except that he has swapped places with Liz, and let him say what he prescribes for that situation. Insofar as John is a good commonsense moralist, he will prescribe that the wife send me a postcard, even though that is (causally) inconsistent with the husband sending me one. For in the imaginary situation about which John is now prescribing, he *is* the wife, and he must properly regard the keeping of his promises as more important. So (PK) does not pass the test; preferences that underlie it are not universalizable. And this should come as no surprise, since, on the present account, the preference that underlies (PK) is a preference for a property that is *not* necessarily shared by worldmates. That is just because (PK) is an agent centred norm.

So agent centred norms are not universalizable, in Hare's sense. Hare would no doubt say that this shows what is wrong with such norms. But it might instead show something to be wrong with his form of universalizability. There is another test that agent centred norms do pass, and it may have just as much right to the name as Hare's test.

We may ask whether, when I consider various possible situations, I prefer the same propositions. But we might ask instead whether I prefer the same properties. This test is extremely weak. But taking it does reveal my true colours. If I were an egomaniac, I would prefer the same property regarding each possible situation: the property of being such that Jamie is successful. If I were an egoist, I would prefer constantly, too, but the property would be the property of being successful. Egomania is not formally ruled out, but it seems unlikely to survive critical reflection.²⁶

Brown University

²⁵ This may not be obvious, so let's check. Suppose Josie prefers a property, P, over its complement, and suppose that P is a property that someone in a world W has while someone else in W doesn't have it. Call some bearer of P in W, 'Lucky,' and some non-bearer 'Sad'. Does Josie prefer that Lucky has the property, or that Sad has it? She can't say, until she knows which she is. For by hypothesis, she prefers the property. She prefers to have it, not that it be instantiated by Lucky, or by Sad.

²⁶ We might try to rule it out. We might require that the properties counted in the determination of moral norms include no proper names, so to speak. What that really amounts to is a requirement that the agent rank universal properties, and not singular properties. A universal proposition is one whose truth requires the existence of no one in particular; likewise, a universal property can be instantiated notwithstanding the existence of anyone in particular. So we might try to argue for a formal restriction on moral norms in this way; but for my part, I see no particular advantage to any such restriction.

REFERENCES

1. S. Blackburn, *Spreading the Word* (Oxford: Blackwell, 1984).
2. J. Dreier, 'Internalism and Speaker Relativism', *Ethics* 101 (1990) pp. 6-26.
3. R. Firth, 'Ethical Absolutism and the Ideal Observer', in J. Hospers, and W. Sellars (ed.), *Readings in Ethical Theory* (Englewood Cliffs, NJ: Prentice-Hall, 1970) pp. 200-221.
4. A. Gibbard, 'Moral Concepts: Substance and Sentiment', in J. Tomberlin (ed.), *Philosophical Perspectives 6: Ethics* (Atascadero, California: Ridgeview Publishing Company, 1992) pp. 199-221.
5. R. M. Hare, *The Language of Morals* (Oxford: Clarendon Press, 1952).
6. R. M. Hare, *Moral Thinking: Its Levels, Method, and Point* (Oxford: Clarendon Press, 1981).
7. I. L. Humberstone, 'Two Kinds of Agent-Relativity', *Philosophical Quarterly* 41 (1991) pp. 144-166.
8. S. Kagan, *The Limits of Morality* (New York: Clarendon Press, 1989).
9. J. Kalin, 'In Defense of Egoism', in D. Gauthier (ed.), *Morality and Self-Interest* (Englewood Cliffs: Prentice-Hall, 1970).
10. D. Kaplan, 'Demonstratives', in J. Almog, J. Perry and H. Wettstein (ed.), *Themes from Kaplan* (New York: Oxford University Press, 1989) pp. 481-563.
11. D. K. Lewis, 'Attitudes *De Dicto* and *De Se*', in D. K. Lewis (ed.), *Philosophical Papers* (New York: Oxford University Press, 1983) pp. 133-156.
12. B. Medlin, 'Ultimate Principles and Ethical Egoism', *Australasian Journal of Philosophy* 35 (1957) pp. 111-118.
13. T. Nagel, *The Possibility of Altruism* (Princeton: Princeton University Press, 1970).
14. T. Nagel, *The View from Nowhere* (New York: Oxford University Press, 1986).
15. J. Perry, 'The Problem of the Essential Indexical', *Nous* 13 (1979) pp. 3-21.
16. S. Scheffler, *The Rejection of Consequentialism* (New York: Clarendon Press, 1982).
17. A. Sen, 'Rights and Agency', *Philosophy and Public Affairs* 11 (1982) pp. 3-29.
18. A. Sen, 'Evaluator Relativity and Consequential Evaluation', *Philosophy and Public Affairs* 12 (1983) pp. 113-32