Roughly, values are what we are disposed to value. Less roughly, we have this schematic definition: *Something of the appropriate category is a value if and only if we would be disposed, under ideal conditions, to value it.* It raises five questions. (1) What is the favourable attitude of ‘valuing’? (2) What is the ‘appropriate category’ of things? (3) What conditions are ‘ideal’ for valuing? (4) Who are ‘we’? (5) What is the modal status of the equivalence?

By answering these questions, I shall advance a version of the dispositional theory of value. I begin by classifying the theory that is going to emerge. First, it is naturalistic: it advances an analytic definition of value. It is naturalistic in another sense too: it fits into a naturalistic metaphysics. It invokes only such entities and distinctions as we need to believe in anyway, and needs nothing extra before it can deliver the values. It reduces facts about value to facts about our psychology.

The theory is subjective: it analyses value in terms of our attitudes. But it is not subjective in the narrower sense of implying that value is a topic on which whatever we may think is automatically true, or on which there is no truth at all. Nor does it imply that if we had been differently disposed, different things would have been values. Not quite—but it comes too close for comfort.

The theory is internalist: it makes a conceptual connection between value and motivation. But it offers no guarantee that everyone must be motivated to pursue whatever is of value; still less, whatever he judges to be of value. The connection is defeasible, in more ways than one.

The theory is cognitive: it allows us to seek and to gain knowledge about what is valuable. This knowledge is *a posteriori* knowledge of contingent matters of fact. It could in principle be gained by psychological experimentation. But it is more likely to be gained by difficult exercises of imagination, carried out perhaps in a philosopher’s or a novelist’s armchair.
The theory is conditionally relativist: it does not exclude the possibility that there may be no such thing as value *simpliciter*, just value for this or that population. But it does not imply relativity, not even when taken together with what we know about the diversity of what people actually value. It leaves the question open.

Is it a form of realism about value?—That question is hard. I leave it for the end.

*What is 'valuing'?* It is some sort of mental state, directed toward that which is valued. It might be a feeling, or a belief, or a desire. (Or a combination of these; or something that is two or three of them at once; or some fourth thing. But let us set these hypotheses aside, and hope to get by with something simpler.)

A feeling?—Evidently not, because the feelings we have when we value things are too diverse.

A belief? What belief? You might say that one values something just by believing it to be a value. That is circular. We might hide the circularity by maneuvering between near-synonyms, but it is better to face it at once. If so, we have that being a value is some property such that something has it iff we are disposed, under ideal conditions, to believe that the thing has it. In other words, such that we are disposed, under ideal

---

1 The most interesting of the hypotheses here set aside is that an attitude of valuing might be a ‘desire’; a special kind of attitude that is both a belief and a desire and that motivates us, without benefit of other desires, in just the way that ordinary desires do. (Or it might be an attitude that is not identical with, but rather is necessarily connected with, a belief and a desire; or an attitude that is not strictly speaking either a belief or a desire, but is just like each apart from also being like the other.) Valuing X might be the desire that is at once a belief that X is good and a desire for X; where *goodness* just means that property, whatever it may be, such that a belief that X has it may double as a desire for X.

But we should hesitate to believe in besires, because integrating them into the folk psychology of belief and desire turns out to be no easy thing. On the difficulty with instrumental besires, see my ‘Desire as Belief’ and John Collins, ‘Belief, Desire and Revision’, *Mind* 97 (1988), pp. 323–342: when a system of attitudes changes under the impact of new information, beliefs evolve in one way and (instrumental) desires in another. A desire, trying to go both ways at once, would be torn apart. Intrinsic besires—a better candidate for the attitude of valuing—face a different difficulty. At least in miniature examples, they turn out to be altogether impervious to change under the impact of experience. Not bad, you might think—why should experience change our mind about what’s intrinsically good? The trouble is that the result applies not only to perceptual experience but also to experience of moral reflection, ‘intuiting’, and the like.
conditions, to be right about whether something has it. That is not empty; but it tells us little, since doubtless there are many properties about which we are disposed to be right.

Further, if valuing something just meant having a certain belief about it, then it seems that there would be no conceptual reason why valuing is a favourable attitude. We might not have favoured the things we value. We might have opposed them, or been entirely indifferent.

So we turn to desires. But we'd better not say that valuing something is just the same as desiring it.\(^2\) That may do for some of us: those who manage, by strength of will or by good luck, to desire exactly as they desire to desire. But not all of us are so fortunate. The thoughtful addict may desire his euphoric daze, but not value it. Even apart from all the costs and risks, he may hate himself for desiring something he values not at all. It is a desire he wants very much to be rid of.\(^3\) He desires his high, but he does not desire to desire it, and in fact he desires not to desire it. He does not desire an unaltered, mundane state of consciousness, but he does desire to desire it. We conclude that he does not value what he desires, but rather he values what he desires to desire.

Can we do better by climbing the ladder to desires of even-higher order? What someone desires to desire to desire might conceivably differ from what he does desire to desire. Or. . . . Should we perhaps say that what a person really values is given by his highest order of desire, whatever order that is?—It is hard to tell whether this would really be better, because it is hard to imagine proper test cases.\(^4\) Further, if we go for the highest

---

\(^2\) Often in decision theory and economics, 'value' does just mean a measure of desiredness, and all desires count equally. But it's not the sense we want here.


\(^4\) It is comparatively easy to imagine instrumental third-order desires. Maybe our addict wishes he could like himself better than he does; and not by doing away with his addiction, which he takes to be impossible, but by becoming reconciled to it and accepting himself as he is. Or maybe he just fears that his second-order desire not to be addicted will someday lead him to suffer the pains of withdrawal. Either way, he wants to be rid of his second-order desire not to be addicted, but he wants it not for itself but as a means to some end. This is irrelevant: presumably it is intrinsic, not instrumental, desiring that is relevant to what someone values.
order, we automatically rule out the case of someone who desires to value differently than he does, yet this case is not obviously impossible. I hesitantly conclude we do better to stop on the second rung: valuing is just desiring to desire.

Recall G. E. Moore: 'To take, for instance, one of the more plausible, because one of the more complicated, of such proposed definitions, it may easily be thought, at first sight, that to be good may mean to be that which we desire to desire.'\(^5\) Of course he does not endorse the definition, but at least he does it the honour of choosing it for his target to display the open question argument. I don't say that everything we value is good; but I do echo Moore to this extent. I say that to be valued by us means to be that which we desire to desire. Then to be a value—to be good, near enough—means to be that which we are disposed, under ideal conditions, to desire to desire. Still more complicated, still more plausible. It allows, as it should, that under less-than-ideal conditions we may wrongly value what is not really good. As for Moore's open question, we shall face that later.

We have this much of an 'internalist' conceptual connection between value and motivation. If something is a value, and if someone is one of the appropriate 'we', and if he is in ideal conditions, then it follows that he will value it. And if he values it, and if he desires as he desires to desire, then he will desire it. And if he desires it, and if this desire is not outweighed by other conflicting desires, and if he has the instrumental rationality to do what serves his desires according to his beliefs, then he will pursue it. And if the relevant beliefs are near enough true, then he will pursue it as effectively as possible. A conceptual connection between value and motivation, sure enough—but a multifariously iffy connection. Nothing less iffy would be credible. But still less is it credible that there is no connection at all.

In general, to find out whether something is disposed to give response R under conditions C, you can put it in C and find out whether you get R. That is a canonical way to learn whether the disposition is present, though surely not the only possible

\(^5\) *Principia Ethica* (Cambridge University Press, 1903) Section 13.
way. If a dispositional theory of value is true, then we have a canonical way to find out whether something is a value. To find out whether we would be disposed, under ideal conditions, to value it, put yourself in ideal conditions, if you can, making sure you can tell when you have succeeded in doing so. Then find out whether you value the thing in question, i.e. whether you desire to desire it. If you do, that confirms that it is a value. (I assume you are one of the appropriate ‘we’ and you know it.) Now we have this much of an ‘internalist’ conceptual connection between value judgements and motivation. It is even iffier than the connection between value itself and motivation; and again I say that if it were less iffy, it would be less credible. If someone believes that something is a value, and if he has come to this belief by the canonical method, and if he has remained in ideal conditions afterward or else retained the desire to desire that he had when in ideal conditions, then it follows that he values that thing. And if he desires as he desires to desire, then he desires that thing; and so on as before.

The connection is not with the judgement of value per se, but with the canonical way of coming to it. If someone reached the same judgement in some non-canonical way—as he might—-that would imply nothing about his valuing or desiring or pursuing.

What is the ‘appropriate category’? If values are what we are disposed to desire, then the things that can be values must be among the things that can be desired. Those fall into two classes. Sometimes, what one desires is that the world should be a

---

6 It is a fallible way; for it may be that you cannot put the thing in C without making the disposition disappear. Imagine that a surface now has just the molecular structure that disposes things to reflect light; but that exposing it to light would catalyze a swift chemical change and turn it into something unreflective. So long as it’s kept in the dark, is it reflective?—I think so; but its reflectivity is what Ian Hunt once called a ‘finkish’ disposition, one that would vanish if put to the test. (So a simple counterfactual analysis of dispositions fails.) Could a disposition to value, or to disvalue, be finkish? Yes; here is an example due to Michael Tooley. Suppose, as I shall claim, that ‘ideal conditions’ include imaginative acquaintance; suppose there is no way to imagine direct electrical stimulation of the pleasure centre of the brain except by trying it out; and suppose that one brief trial would enslave you to the electrode and erase all other desires. Then I think you might well have a finkish disposition to disvalue the experience. If, per impossibile, you could manage to imagine it without at the same time having your present system of desires erased by the current, you would desire not to desire it.
certain way: that it should realise one of a certain class of (maximally specific, qualitatively delineated) possibilities for the whole world. This class—a ‘proposition’, in one sense of that word—gives the content of the desire. To desire that the world realise some possibility within the class is to desire that the proposition be true. Call this ‘desire \textit{de dicto}'.

But sometimes, what one desires concerns not just the world but oneself: one simply desires to \textit{be} a certain way. For instance, Fred might want to be healthy, or wealthy, or wise. Then what he wants is that he himself should realise one of a certain class of (maximally specific, qualitatively delineated) possibilities for an individual—or better, for an individual-in-a-world-at-a-time. This class—a ‘property’ in one sense of that word, or an ‘egocentric proposition’—gives the content of the desire. To desire to realise some possibility in the class is to desire to have the property, or to desire that the egocentric proposition be true of one. Call this ‘desire \textit{de se}', or ‘egocentric’ or ‘essentially indexical’ desire.\footnote{See Peter Geach, ‘On Beliefs about Oneself’, \textit{Analysis} 18 (1957), pp. 23–24; Hector-Neri Castañeda ‘On the Logic of Attributions of Self-Knowledge to Others’, \textit{Journal of Philosophy} 65 (1968), pp. 439–456; John Perry, ‘Frege on Demonstratives’, \textit{Philosophical Review} 86 (1977), pp. 474–497, and ‘The Problem of the Essential Indexical’, \textit{Noûs} 14 (1979), pp. 3–21; my ‘Attitudes \textit{De Dicto} and \textit{De Se}', \textit{Philosophical Review} 88 (1979), pp. 513–543; Roderick Chisholm, \textit{The First Person: An Essay on Reference and Intentionality} (Harvester Press, 1981).}

You might think to reduce desire \textit{de se} to desire \textit{de dicto}, saying that if Arthur desires to be happy, what he desires is that the world be such that Arthur is happy. (You might doubt that such worlds comprise a qualitatively delineated class, so you might consider dropping that requirement.) But no. That is not exactly the same thing, though the difference shows up only when we imagine someone who is wrong or unsure about who in the world he is. Suppose Arthur thinks he is Martha. If Arthur is self-centred he may desire to be happy, desire that the world be one wherein Martha is happy, but not desire that the world is one wherein Arthur is happy. If instead Arthur is selflessly benevolent he may not desire to be happy, yet he may desire that the world be such that Arthur is happy. If Arthur is so befuddled as not to know whether he is Arthur or Martha, but hopes he is Arthur, he does not just desire that the world be such that
Arthur is self-identical! In all these cases, Arthur’s desire is, at least in part, irreducibly de se.\(^8\)\(^9\) When we acknowledge desires de se, we must distinguish two senses of ‘desiring the same thing’. If Jack Sprat and his wife both prefer fat meat, they desire alike. They are psychological duplicates, on this matter at least. But they do not agree in their desires, because no possible arrangement could satisfy them both. Whereas if Jack prefers the fat and his wife prefers the lean, then they differ psychologically, they do not desire alike. But they do agree, because if he eats no fat and she eats no lean, that would satisfy them both. In general, they desire alike iff they desire de se to have exactly the same properties and they desire de dicto that exactly the same propositions hold. They agree in desires iff exactly the same world would satisfy the desires of both; and a world that satisfies someone’s desires is one wherein he has all the properties that he desires de se and wherein all the propositions hold that he desires de dicto. Agreement in desire makes for harmony; desiring alike may well make for strife.

As we can desire de dicto or de se, so we can desire to desire de dicto or de se. If desiring to desire is valuing, and if values are what we are disposed to value, then we must distinguish values de dicto and de se. A value de dicto is a proposition such that we are disposed to desire to desire de dicto that it hold. A value de se is a property such that we are disposed to desire to desire de se to have it.

It is essential to distinguish. Consider egoism: roughly, the thesis that one’s own happiness is the only value. Egoism is meant to be general. It is not the thesis that the happiness of a certain special person, say Thrasymachus, is the only value. Egoism de dicto says that for each person X, the proposition that

\(^8\)What we can do is to go the other way, subsuming desire de dicto under desire de se. To desire that the world be a certain way is to desire that one have the property of living in a world that is that way—a property that belongs to all or none of the inhabitants of the world, depending on the way the world is. This subsumption, artificial though it be, is legitimate given a suitably broad notion of property. But for present purposes we need distinction, not unification. So let us henceforth ignore those desires de se that are equivalent to desires de dicto, and reserve the term ‘de se’ for those that are not.

\(^9\)If you like, you can put the egocentricity not in the content of desire itself but in an egocentric mode of presentation of that content. The choice matters little, save to simplicity. See Jeremy Butterfield, ‘Content and Context’ in Butterfield, ed., Language Mind and Logic (Cambridge University Press, 1986).
X is happy is the only value. That is inconsistent, as Moore observed. It says that there are as many different values as there are people, and each of them is the only value. Egoism *de se* says that the property of happiness—in other words, the egocentric proposition that one is happy—is the only value. Moore did not confute that. He ignored it. False and ugly though it be, egoism *de se* is at least a consistent doctrine. What it alleges to be the only value would indeed be just one value *de se*, not a multitude of values *de dicto*.

Insofar as values are *de se*, the wholehearted pursuit by everyone of the same genuine value will not necessarily result in harmony. All might value alike, valuing *de se* the same properties and valuing *de dicto* the same propositions. Insofar as they succeed in desiring as they desire to desire, they will desire alike. But that does not ensure that they will agree in desire. If egoism *de se* were true, and if happiness could best be pursued by doing others down and winning extra shares, then the pursuit by all of the very same single value would be the war of all against all.

Because egoism is false and ugly, we might be glad of a theoretical framework that allowed us to confute it *a priori*. And some of us might welcome a framework that promises us harmony, if only we can all manage to pursue the same genuine values. Was it right, then, to make a place for values *de se*? Should we have stipulated, instead, that something we are disposed to desire to desire shall count as a value only when it is a proposition that we are disposed to desire to desire *de dicto*?

No. Probably it is already wrong to reject egoism *a priori* but, be that as it may, there are other doctrines of value *de se*, more plausible and more attractive. Self-improvement and self-sacrifice are no less egocentric than self-aggrandizement and

---

10 *Principia Ethica*, Section 59.

11 Someone who said that happiness was the only value might mean something else, which is not a form of egoism at all. He might mean that the proposition that happiness is maximized is the only value—a single value *de dicto*. Or he might mean that for each person X, the proposition that X is happy is a value *de dicto*, and that these many values of parallel form are the only values. Mean what you please—I take these to be legitimate, but derivative, senses in which a property may be called a value. I only say they should not be confused with, or drive out, the sense in which a property may be a value *de se*. 
self-indulgence. Surely we should make a place for putative values *de se* of altruism, of honour, and of loyalty to family, friends, and country.\textsuperscript{12} We may entertain the substantive thesis that none of these putative values *de se* is genuine, and that all genuine values are *de dicto*. But even if we believed this—myself, I think it wildly unlikely—we should not beg the question in its favour by building it into our theoretical framework.

*What conditions are ‘ideal’?* If someone has little notion what it would be like to live as a free spirit unbound by law, custom, loyalty, or love; or what a world of complete harmony and constant agreement would be like; then whether or not he blindly values these things must have little to do with whether or not they are truly values. What he lacks is imaginative acquaintance. If only he would think harder, and imagine vividly and thoroughly how it would be if these putative values were realised (and perhaps also how it would be if they were not) that would make his valuing a more reliable indicator of genuine value. And if he could gain the fullest imaginative acquaintance that is humanly possible,\textsuperscript{13} then, I suggest, his valuing would be an infallible indicator. Something is a value iff we are disposed, under conditions of the fullest possible imaginative acquaintance, to value it.

Compare a version of Intuitionism: by hard thought, one becomes imaginatively well acquainted with X; in consequence, but not as the conclusion of any sort of inference, one intuits that X has a certain unanalysable, non-natural property; and in consequence of that, one comes to value X. My story begins and ends the same. Only the middle is missing. Again, an exercise of imaginative reason plays a crucial role. Again, its relation to what follows is causal, and in no way inferential. But in my story, the consequent valuing is caused more directly, not via the detection of a peculiar property of X.

Can we say that the valuing ensued because X was a value?—Maybe so, but if we do, we are not saying much: it


\textsuperscript{13} Without in the process having his dispositions to value altered—see Footnote 6.
ensues because there is something about imaginative acquaintance with X that causes valuing.\textsuperscript{14}

The canonical way to find out whether something is a value requires a difficult imaginative exercise. And if you are to be sure of your answer, you need to be sure that you have gained the fullest imaginative acquaintance that is humanly possible. A tall order! You had better settle for less. Approximate the canonical test. Try hard to imagine how it would be if the putative value were (or were not) realised. Hope that your acquaintance comes close enough to the fullest possible that getting closer would not change your response. Then you may take your valuing as fallible evidence that you were acquainted with a genuine value, or your indifference as fallible evidence that you were not. You cannot be perfectly certain of your answer, but you can take it as sure enough to be going on with, subject to reconsideration in the light of new evidence. How sure is that?—Well, as always when we acknowledge fallibility, some of us will be bolder than others.

New evidence might be a more adequate imaginative exercise of your own. It might be the testimony of others. It might in principle be a result of scientific psychology—though it is far from likely that any such results will come to hand soon!

A trajectory toward fuller imaginative acquaintance with putative value X is not just a sequence of changes in your imaginative state. It has a direction to it. And that is so independently of my claim that it leads, after a point, to ever-surer knowledge about whether X is a value. For in learning how to imagine X, you gain abilities; later you have all the relevant imaginative abilities you had before, and more besides. And you notice, \textit{a priori}, relationships of coherence or incoherence between attitudes that might figure in the realisation of X; later you are aware of all that you had noticed before, and more

\textsuperscript{14}How does imaginative acquaintance cause valuing, when it does? How does imagination render values attractive? Does it happen the same way for all values?—For our purposes, it is enough to say that it happens. We needn't know how. But we may guess. Maybe imaginative acquaintance shows us how new desires would be seamless extensions of desires we have already. Or maybe we gravitate toward what we understand, lest we baffle ourselves—see J. David Velleman, \textit{Practical Reflection} (Princeton University Press, forthcoming). But that cannot be the whole story, because some easily understood lives—say a life of lethargy, ruled by a principle of least action—remain repellent.
besides. And you think of new questions to explore in your imagining—what might the life of the free spirit become, long years after its novelty had worn off?—and later you have in mind all the questions you had thought of before, and more besides. Forgetting is possible, of course. But by and large, the process resists reversal.  

Our theory makes a place for truth, and in principle for certain knowledge, and in practice for less-than-certain knowledge, about value. But also it makes a place for ignorance and error, for hesitant opinion and modesty, for trying to learn more and hoping to succeed. That is all to the good. One fault of some subjective and prescriptive theories is that they leave no room for modesty: just decide where you stand, then you may judge of value with the utmost confidence!

There is a long history of theories that analyse value in terms of hypothetical response under ideal conditions, with various suggestions about what conditions are ideal. Imaginative acquaintance often gets a mention. But much else does too. I think imaginative acquaintance is all we need—the rest should be in part subsumed, in part rejected.

First, the responder is often called an ideal spectator. That is tantamount to saying that conditions are ideal only when he is observing a sample of the putative value in question (or of its absence). If the putative value is de se, a property, then a sample can just be an instance. If it is de dicto, a proposition, it is hard to say in general what an observable sample could be. But if it is the proposition that a certain property is instantiated sometimes, or often, or as often as possible, or in all cases of a certain kind, then again a sample can just be an instance of the property. Anyone happy may serve as a sample of the proposition that total happiness is maximised.

Observable samples can sometimes prompt the imagination and thereby help us to advance imaginative acquaintance. But they are of limited use. For one thing, observation does not include mind-reading. Also, it does best with short, dramatic episodes. A lifelong pattern of stagnation, exemplifying the absence of various values, goes on too long to be easily

---

observed. Samples are dispensable as aids to imagination, and sometimes they are comparatively ineffective. A novel might be better.

The notion of an ideal spectator is part of a longstanding attempt to make dispositional theories of value and of colour run in parallel. But the analogy is none too good, and I doubt that it improves our understanding either of colour or of value. Drop it, and I think we have no further reason to say that a disposition to value is a disposition to respond to observed samples. 16

Second, the ideal responder is often supposed to be well informed. If any item of empirical knowledge would affect his response, he knows it.—But some sorts of knowledge would not help to make your valuing a more reliable indicator of genuine value. Instead they would distract. If you knew too well how costly or how difficult it was to pursue some value, you might reject the grapes as sour, even when imaginative acquaintance with the value itself would have caused you to value it. Genuine values might be unattainable, or unattainable without undue sacrifice of other values. An ideal balancer of values needs thorough knowledge of the terms of trade. An ideal valuer may be better off without it. Our present business is not with the balancing, but with the prior question of what values there are to balance. 17

Another unhelpful sort of knowledge is a vivid awareness that we are small and the cosmos is large; or a vivid awareness of the mortality of mankind, and of the cosmos itself. If such knowledge tends to extinguish all desire, and therefore all

16 If we had demanded samples, we would have had a choice about where to locate the disposition. Is it within us or without? Is it a disposition in the samples to evoke a response from spectators?—that is what best fits the supposed parallel with a dispositional theory of colour. See Robert Pargetter and John Campbell, 'Goodness and Fragility', American Philosophical Quarterly 23 (1986), pp. 155-166, for an analysis of this kind. Or is it a disposition in the spectators to respond to samples? Or is it a disposition of the sample-cum-spectator system to respond to having its parts brought together? For us there is no choice. The propositions and properties that are the values cannot harbour any causal bases for dispositions. Samples could, but there needn't be any samples. Imaginative experiences could, but those are within us, and are not themselves samples of values. So the disposition must reside in us, the responders. Being a value comes out as a dispositionally analysed property, but not as a disposition of the things that have it. Values themselves are not disposed to do anything.

17 Previous theories of hypothetical response may indeed have been concerned as much with the analysis of right balancing as with value itself. If so, they cannot be faulted for trying to characterise an ideal balancer. However my present analysandum is different.
VALUING, IT WILL NOT HELP US TO VALUE JUST WHAT IS VALUABLE. LIKEWISE IT WILL BE UNHELPFUL TO DWELL TOO MUCH ON THE LOWLY CAUSAL ORIGINS OF THINGS. IF SOME FEATURE OF OUR LIVES ORIGINATED BY KIN SELECTION, OR PAVLOVIAN CONDITIONING, OR SUBLIMATION OF INFANTILE SEXUALITY, THAT IS IRRELEVANT TO WHAT IT IS LIKE IN ITSELF. UNLESS HE CAN OVERCOME THE ILLUSION OF RELEVANCE, A VALUER WILL BE MORE RELIABLE IF HE REMAINS IGNORANT OF SUCH MATTERS.

HOWEVER, I GRANT ONE CASE—A COMMON ONE—IN WHICH ONE DOES NEED EMPIRICAL KNOWLEDGE IN ORDER TO GAIN IMAGINATIVE ACQUAINTANCE WITH A GIVEN PUTATIVE VALUE. IT MAY BE ‘GIVEN’ IN A WAY THAT UNDERSPECS IT, WITH THE REST OF THE SPECIFICATION LEFT TO BE FILLED IN BY REFERENCE TO THE ACTUAL WAYS OF THE WORLD. FOR INSTANCE WHEN I MENTIONED THE LIFE OF A FREE SPIRIT AS A PUTATIVE VALUE, WHAT I MEANT—AND WHAT YOU SURELY TOOK ME TO MEAN—WAS THE LIFE OF A FREE SPIRIT IN A WORLD LIKE OURS. IN SUCH CASES, A VALUER MUST COMPLETE THE SPECIFICATION BY DRAWING ON HIS KNOWLEDGE OF THE WORLD, ELSE HE WILL NOT KNOW WHAT HE IS SUPPOSED TO IMAGINE. TO THAT EXTENT—and only to that extent, I THINK—BEING WELL-INFORMED IS INDEED A QUALIFICATION FOR HIS JOB.

THIRD, IT MAY BE SAID THAT THE IDEAL RESPONDER SHOULD NOT ONLY IMAGINE HAVING (OR LACKING) A PUTATIVE VALUE, BUT ALSO IMAGINE THE EFFECT ON OTHER PEOPLE OF SOMEONE’S HAVING (OR LACKING) IT. THINKING WHAT IT WOULD BE LIKE TO LIVE AS A FREE SPIRIT IS NOT ENOUGH. YOU MUST ALSO THINK WHAT IT WOULD BE LIKE TO ENCOUNTER THE FREE SPIRIT AND BE ILL-USED.—BUT AGAIN, I THINK THE REQUIREMENT IS MISPLACED. IT IS APPROPRIATE NOT TO AN IDEAL VALUER, BUT TO AN IDEAL BALANCER WHO MUST THINK THROUGH THE COST TO SOME VALUES OF THE REALISATION OF OTHERS. IN ADDRESSING THE PRIOR QUESTION OF WHAT VALUES THERE ARE, COUNTING THE COST IS A DISTRACTION TO BE RESISTED.

OFTEN, HOWEVER, REALISING A PUTATIVE VALUE DE SE WOULD ITSELF INVOLVE IMAGINING THE IMPACT OF ONE’S CONDUCT ON OTHER PEOPLE. WHEN THAT IS SO, IMAGINING REALISING THE VALUE INVOLVES

---

18 IMAGINATIVE ACQUAINTANCE IS SOMETIMES THOUGHT TO CONSIST IN THE POSSESSION OF A SPECIAL KIND OF ‘PHENOMENAL’ INFORMATION. IF THAT IS SO, OF COURSE MY OWN CANDIDATE FOR ‘IDEAL CONDITIONS’ COMES DOWN TO A SPECIAL CASE OF BEING WELL-INFORMED. BUT IT IS NOT SO—not even in the most favourable case, that of imaginative acquaintance with a kind of sense-experience. See my ‘What Experience Teaches’ in William Lycan, ed., Mind and Cognition: A Reader (Blackwell, 1989).

This content downloaded from 130.132.173.105 on Tue, 4 Jun 2013 14:06:34 PM
All use subject to JSTOR Terms and Conditions
imagining the impact; and that cannot be done without simply imagining the impact. In such cases, imagining the impact does fit in; for it is already subsumed as part of imaginative acquaintance with the value itself.

Fourth, the ideal responder is often said to be dispassionate and impartial, like a good judge.—Once more, the requirement is appropriate not to an ideal valuer but to an ideal balancer. The valuer is not a judge. He is more like an advocate under the adversarial system. He is a specialist, passionate and partial perhaps, in some one of all the values there are. On the present theory, when I say that X is a value iff we are disposed to value X under ideal conditions, I do not mean conditions that are ideal *simpliciter*, but rather conditions that are ideal *for* X. We should not assume that there is any such thing as a condition of imaginative acquaintance with all values at once. (Still less, all putative values.) Imagination involves simulation—getting into the skin of the part. How many skins can you get into all at once? Tranquillity and vigorous activity might both be values; but a full imaginative acquaintance with one might preclude a full imaginative acquaintance with the other. (The incompatibility might even be conceptual, not just psychological.) Then if we value both, as surely many of us do, it is not because of acquaintance with both at once. It might be a lasting effect of past imaginative acquaintance at some times with one and at other times with the other.

A further speculation: it might happen that there were values that could not even be valued all at once. If so, then conflict of values would go deeper than is ever seen in hard choices; because what makes a choice hard is that conflicting values are valued together by the unfortunate chooser. An alarming prospect!—or exhilarating, to those of us who delight in the rich variety of life.

*Who are ‘we’?* An *absolute* version of the dispositional theory says that the ‘we’ refers to all mankind. To call something a value is to call it a value *simpliciter*, which means that everyone, always and everywhere, is disposed under ideal conditions to value it. Then there are values only insofar as all mankind are alike in their dispositions.

Maybe all mankind *are* alike. The manifest diversity of
valuing between different cultures—or for that matter within a culture, say between colleagues in the same philosophy department—is no counterevidence. In the first place, people may not be valuing as they would be disposed to value under ideal conditions. In the second place, remember that conditions of imaginative acquaintance are ideal for particular values, not *simpliciter*. So even if all are disposed alike, and all value as they would under ideal conditions, that may mean that some people value X as they would under conditions ideal for X, while others, who are no differently disposed, value Y as they would under conditions ideal for Y. If no conditions are ideal at once for X and for Y (still more if X and Y cannot both be valued at once), there could be diversity of valuing even in a population of psychological clones, if different ones had been led into different imaginative exercises.

We saw that it would be no easy job to find out for sure whether a particular person would be disposed to value something under ideal conditions of imaginative acquaintance with it. It would be harder still to find out all about one person’s dispositions. And not just because one hard job would have to be done many times over. It might happen that imaginative acquaintance with X would leave traces, in one’s valuing or otherwise, that got in the way of afterward imagining Y. To the extent that there was such interference, each new imaginative experiment would be harder than the ones before.

The fallback, if we are wary of presupposing that all mankind are alike in their dispositions to value, is tacit relativity. A *relative* version says that the ‘we’ in the analysis is indexical, and refers to a population consisting of the speaker and those somehow like him. If the analysis is indexical, so is the analysandum. Then for speaker S to call something a value is to call it a value for the population of S and those like him; which means that S and those like him are all disposed, under ideal conditions, to value it.

The relative version is not just one version, but a spectrum. What analysis you get depends on how stringent a standard of similarity you apply to the phrase ‘the speaker and those somehow like him’. At one end of the spectrum stands the absolute version: common humanity is likeness enough, so whoever speaks, all mankind are ‘we’. At the other end, ‘we’
means: ‘you and I, and I’m none too sure about you’. (Or it might be ‘I, and those who think as I do’, which reduces to ‘I’.) In between, ‘we’ means: ‘I, and all those who are of a common culture with me’. Since mankind even at one moment is not made up of isolated and homogeneous tribes, and since we should not limit ourselves to the part of mankind located at one moment, we may haggle endlessly over how much cultural affiliation is meant.

(We have a piece of unfinished business: if someone is to find out about values by the canonical method, he must somehow know that he is one of the appropriate ‘we’. All our versions, absolute or relative, make this knowledge automatic. Not so for elitist versions, on which ‘we’ means ‘the best-qualified of us’ or maybe ‘the most normal of us’. But elitist versions are pointless. We’re already considering dispositions under extravagantly ideal conditions; we needn’t idealise all over again by being selective about who counts as one of the ‘we’.)

If some relative version were the correct analysis, wouldn’t that be manifest whenever people talk about value? Wouldn’t you hear them saying ‘value for me and my mates’ or ‘value for the likes of you’? Wouldn’t you think they’d stop arguing after one speaker says X is a value and the other says it isn’t?—Not necessarily. They might always presuppose, with more or less confidence (well-founded or otherwise), that whatever relativity there is won’t matter in this conversation. Even if they accept in principle that people sometimes just differ in their dispositions to value, they may be very reluctant to think the present deadlocked conversation is a case of such difference. However intractable the disagreement may be, they may go on thinking it really is a disagreement: a case in which two people are disposed alike, but one of them is wrong about what is a value relative to their shared dispositions, because he is not valuing as he would under ideal conditions. So long as they think that—and they might think it very persistently—they can hold the language of explicit relativity in reserve. It is there as a last resort, if ever they meet with a proven case of ultimate difference. But it will not be much heard, since it is a practical impossibility to prove a case. If the language of absolutism prevails, that is not strong evidence against relativity.

(Those who have heard of the relativity of simultaneity do not
manifest this knowledge all the time. They speak as the ignorant
do, and no harm done. They'll resort to the language of
relativity when it matters, say in discussing the exploits of the
interstellar navy.)

Does the language of absolutism prevail? Not really. With
some of us it does. Others of us resort to the language of relativity
at the drop of a hat. Yet this too is poor evidence. The eager
relativists may have been confused by philosophy. For who can
escape it?

So what version should we prefer, absolute or relative?—
Neither; instead, I commend a **wait-and-see** version. In making a
judgement of value, one makes many claims at once, some
stronger than others, some less confidently than others, and
waits to see which can be made to stick. I say X is a value; I mean
that all mankind are disposed to value X; or anyway all
nowadays are; or anyway all nowadays are except maybe some
peculiar people on distant islands; or anyway . . . ; or anyway
you and I, talking here and now, are; or anyway I am.\(^{19}\) How
much am I claiming?—as much as I can get away with. If my
stronger claims were proven false—though how that could be
proven is hard to guess—I still mean to stand by the weaker
ones. So long as I'm not challenged, there’s no need to back
down in advance; and there’s no need to decide how far I’d back
down if pressed. What I mean to commit myself to is **conditionally
relative**: relative if need be, but absolute otherwise.

What is the modal status of the equivalence? The equivalence
between value and what we are disposed to value is meant to be
a piece of philosophical analysis, therefore analytic. But of
course it is not obviously analytic; it is not even obviously true.

It is a philosophical problem how there can ever be unobvious
analyticity. We need not solve that problem; suffice it to say that
it is everybody’s problem, and it is not to be solved by denying
the phenomenon. There are perfectly clear examples of it: the
epsilon-delta analysis of an instantaneous rate of change, for
one. Whenever it is analytic that all A’s are B’s, but not
obviously analytic, the Moorean open question—whether all
A’s are indeed B’s—is intelligible. And not only is it intelligible

---

\(^{19}\)See the discussion of ‘anyway’ in Frank Jackson, ‘On Assertion and Indicative
in the sense that we can parse and interpret it (that much is true even of the question whether all A’s are A’s) but also in the sense that it makes sense as something to say in a serious discussion, as an expression of genuine doubt.

Besides unobvious analyticity, there is equivocal analyticity. Something may be analytic under one disambiguation but not another, or under one precisification but not another. Examples abound. Quine was wrong that analyticity was unintelligible, right to doubt that we have many clear-cut cases of it. If differing versions of a concept (or, if you like, different but very similar concepts) are in circulation under the same name, we will get equivocal analyticity. It is analytic under one disambiguation of ‘dog’ that all dogs are male; under one disambiguation of ‘bitch’ that all bitches are canine. It is analytic under some precisifications of ‘mountain’ that no mountain is less than one kilometre high. When analyticity is equivocal, open questions make good conversational sense: they are invitations to proceed under a disambiguation or precisification that makes the answer to the question not be analytic. By asking whether there are mountains less than one kilometre high, you invite your conversational partners to join you in considering the question under a precisification of ‘mountain’ broad enough to make it interesting; yet it was analytic under another precisification that the answer was ‘no’. So even if all is obvious, open questions show at worst that the alleged analyticity is equivocal.

I suggest that the dispositional theory of value, in the version I have put forward, is equivocally as well as unobviously analytic. I do not claim to have captured the one precise sense that the word ‘value’ bears in the pure speech, uncorrupted by philosophy, that is heard on the Clapham omnibus. So far as this matter goes, I doubt that speakers untouched by philosophy are found in Clapham or anywhere else. And if they were, I doubt if they’d have made up their minds exactly what to mean any more than the rest of us have. I take it, rather, that the word ‘value’, like many others, exhibits both semantic variation and semantic indecision. The best I can hope for is that my dispositional theory lands somewhere near the middle of the

---

range of variation and indecision—and also gives something that I, and many more besides, could be content to adopt as our official definition of the word 'value', in the unlikely event that we needed an official definition.

I've left some questions less than conclusively settled: the matter of absolute versus relative versus wait-and-see versions, the details of 'ideal conditions', the question of admitting values de se, the definition of valuing as second-order versus highest-order intrinsic desiring. It would not surprise or disturb me to think that my answers to those questions are only equivocally analytic—but somewhere fairly central within the range of variation and indecision—and that the same could be said of rival answers. Even if no version of the dispositional theory is unequivocally analytic, still it's fair to hope that some not-too-miscellaneous disjunction of versions comes out analytic under most reasonable resolutions of indeterminacy (under some reasonable precisification of 'most' and 'reasonable'.)

If the dispositional theory is only unobviously and equivocally analytic, why think that it's analytic at all?—Because that hypothesis fits our practice. (The practice of many of us, much of the time.) It does seem that if we try to find out whether something is a genuine value, we do try to follow—or rather, approximate—the canonical method. We gain the best imaginative acquaintance we can, and see if we then desire to desire it. In investigating values by the canonical method, we ignore any alleged possibility that values differ from what we're disposed to value. The dispositional theory explains nicely why we ignore it: no such possibility exists.

Now this should sound an alarm. Phenomenalism, behaviourism, and the like might be supported in exactly the same way: we ignore the possibility that our method of investigation deceives us radically, and the alleged explanation is that no such possibility exists. But in those cases, we know better. We know how systematic hallucination might deceive its victim about the world around him, and how a clever actor might deceive everyone he meets about his inner life (and, in both cases, how it might be that experience or behaviour would remain deceptive throughout the appropriate range of counterfactual suppositions). And it doesn't just strike us that such deception is possible somehow. Rather, we can imagine just how it might happen. We
can give a story of deception all the detail it takes to make it convincing. So we must confess that our method of gaining knowledge of the outer world and the inner lives does consist in part of ignoring genuine possibilities—possibilities that cannot credibly be denied.

The case of value is different, because the convincing detail cannot be supplied. Yes, you might think that perhaps the genuine values somehow differ from what we are disposed to value, even under ideal conditions. (Charles Pigden has noted that a misanthrope might think it because he thinks mankind is irremediably depraved.) The conjecture is not unthinkable; the dispositional theory is not obviously analytic; counterexamples are not obviously impossible. That is not yet much evidence of possibility. Better evidence would be a detailed story of just how it might happen that something—something specific—is after all a value that we are not disposed to value, or a non-value that we are disposed to value. But I have no idea how to flesh out the story. Without 'corroborative detail', insistence that there exist such possibilities is 'bald and unconvincing'. This time, nothing outweighs the niceness of explaining the ignoring by denying the possibilities allegedly ignored.

But is it realism? Psychology is contingent. Our dispositions to value things might have been otherwise than they actually are. We might have been disposed, under ideal conditions, to value seasickness and petty sleaze above all else. Does the dispositional theory imply that, had we been thus disposed, those things would have been values? That seems wrong.

No: we can take the reference to our dispositions as rigidified. Even speaking within the scope of a counterfactual supposition, the things that count as values are those that we are actually disposed to value, not those we would have valued in the counterfactual situation. No worries—unless seasickness actually is a value, it still wouldn’t have been a value even if we’d been disposed to value it.

This is too swift. The trick of rigidifying seems more to hinder the expression of our worry than to make it go away. It can still be expressed as follows. We might have been disposed to value seasickness and petty sleaze, and yet we might have been no different in how we used the word ‘value’. The reference of ‘our
actual dispositions’ would have been fixed on different dispositions, of course, but our way of fixing the reference would have been no different. In one good sense—though not the only sense—we would have meant by ‘value’ just what we actually do. And it would have been true for us to say ‘seasickness and petty sleaze are values’.

The contingency of value has not gone away after all; and it may well disturb us. I think it is the only disturbing aspect of the dispositional theory. Conditional relativity may well disturb us too, but that is no separate problem. What comfort would it be if all mankind just happened to be disposed alike? Say, because some strange course of cultural evolution happened to be cut short by famine, or because some mutation of the brain never took place? Since our dispositions to value are contingent, they certainly vary when we take all of mankind into account, all the inhabitants of all the possible worlds. Given the dispositional theory, trans-world relativity is inevitable. The spectre of relativity within our own world is just a vivid reminder of the contingency of value.

If wishes were horses, how would we choose to ride? What would it take to satisfy us? Maybe this new version of the dispositional theory would suit us better: values are what we’re necessarily disposed to value. Then no contingent ‘value’ would deserve the name; and there would be no question of something being a value for some people and not for others, since presumably what’s necessary is a fortiori uniform (unless different dispositions to value are built into different people’s individual essences, an unlikely story).

What kind of necessity should it be? Not mere deontic necessity—values are what we’re disposed to value on pain of being at fault, where the fault in question turns out to consist in failing to be disposed to value the genuine values. That dispositional theory is empty. Its near relatives are nearly empty. And it won’t help to juggle terms; as it might be, by calling it ‘rational necessity’ and then classifying the disposition to value genuine values as a department of ‘rationality’. Probably not nomological necessity either—small comfort to think that we were disposed to disvalue seasickness only because, luckily, our neurons are not subject to a certain fifth force of nature that would distort their workings in just the wrong way. It had better
be necessity *simpliciter*, so-called ‘metaphysical’ necessity.

If we amend the dispositional theory by inserting ‘necessarily’, we can be much more confident that the ‘values’ it defines would fully deserve the name—if there were any of them. But it is hard to see how there possibly could be. If a value, strictly speaking, must be something we are necessarily disposed to value, and if our dispositions to value are in fact contingent, then, strictly speaking, there are no values. If Mackie is right that a value (his term is ‘objective good’) would have to be

sought by anyone who was acquainted with it, not because of any contingent fact that this person, or every person, is so constituted that he desires this end, but just because the end has to-be-pursuedness somehow built into it,

then he is also right to call values ‘queer’ and to repudiate the error of believing in them.21 (Replacing ‘sought’ by ‘valued’ would not change that.) If we amend the dispositional theory, requiring values to be all that we might wish them to be, we bring on the error theory. The fire is worse than the frying pan.

Is it, after all, out of the question that our dispositions to value might be necessary? If the theory of mind I favour is true, then the platitudes of folk psychology do have a certain necessity—albeit conditional necessity—to them.22 There are states that play the functional roles specified in those platitudes, and it is in virtue of doing so that they deserve their folk-psychological names. It is not necessary that there should be any states in us that deserve such names as ‘pain’, ‘belief’, or ‘desire’. But it is necessary that if any states do deserve those names, then they conform to the platitudes. Or rather, they conform well enough. Now suppose that some of the platitudes of folk psychology specified exactly what we were disposed, under ideal conditions, to desire to desire. And suppose those platitudes were non-negotiable: if a system of states did not satisfy them, that would settle that those states did not conform well enough to folk

21 J. L. Mackie, *Ethics: Inventing Right and Wrong* (Penguin, 1977), p. 40. But note that the queerness Mackie has in mind covers more than just the to-know-it-is-to-love-it queerness described in this passage.

psychology to deserve the mental names it implicitly defines. Then there would be things we were necessarily disposed to value—on condition that we had mental lives at all!

The suggestion is intelligible and interesting, but too good to be true. For one thing, it only spreads the trouble. Instead of losing the risk that nothing deserves the name of value, we gain the added risk that nothing deserves commonplace folk-psychological names. *Pace* the Churchlands, it's not really credible that there might turn out to be no beliefs, no desires, no pains, . . . 23 For another thing, it proves too much. It denies outright that it's possible for someone to differ from others in his dispositions to value. Yet this does seem possible; and we can flesh out the story with plenty of 'corroborative detail'. This cunning and subtle villain once was as others are; he gained excellent imaginative acquaintance with many values, and valued them accordingly. Now he has gone wrong, and cares not a fig for what he once valued; and yet he has forgotten nothing. (He certainly has not stopped having any mental life deserving of the name.) He hates those who are as he once was, and outwits them all the better because of his superb empathetic understanding of what they hold dear. Could it not happen?—not if the present suggestion were true. So the present suggestion is false. Yet it was the only hope, or the only one I know, for explaining how there might be things we are necessarily disposed to value. The dispositions are contingent, then. And, at least in some tacit way, we know it. If the story of the subtle villain strikes you as a possible story, that knowledge thereby reveals itself.

But if we know better, it is odd that we are disturbed—as I think many of us will be—by a dispositional theory of value, unamended, according to which values are contingent. It feels wrong. Why might that be?—Perhaps because a large and memorable part of our discussion of values consists of browbeating and being browbeaten. 24 The rhetoric would fall flat if we kept in mind, all the while, that it is contingent how we are disposed to value. So a theory which acknowledges that contingency

23 As argued in Frank Jackson and Philip Pettit, 'In Defence of Folk Psychology', forthcoming in *Philosophical Studies*.

cannot feel quite right. You might say that it is unfaithful to the distinctive phenomenological character of lived evaluative thought. Yet even if it feels not right, it may still be right, or as near right as we can get. It feels not quite right to remember that your friends are big swarms of little particles—it is inadequate to the phenomenology of friendship—but still they are.

I suggested earlier that my version of the dispositional theory of value might be equivocally analytic. So might the amended version, on which values are what we are necessarily disposed to value. Between these two versions, not to mention others, there might be both semantic variation and semantic indecision. If so, it is part of a familiar pattern. One way to create indeterminacy and equivocal analyticity is to define names implicitly in terms of a theory (folk or scientific), and later find out that the theory is wrong enough that nothing perfectly deserves the names so introduced, but right enough that some things, perhaps several rival candidates, deserve the names imperfectly. Nothing perfectly deserves the name ‘simultaneity’, since nothing quite fits the whole of our old conception. So the name will have to go to some imperfect deserter of it, or to nothing. What it takes to deserve this name, not perfectly but well enough, was never officially settled. One resolution of the indeterminacy makes it analytic that simultaneity must be frame-independent; another, that it must be an equivalence relation; a third, that it must be both at once. The third brings with it an error theory of simultaneity.25

I suggest that (for some of us, or some of us sometimes) the amended dispositional theory best captures what it would take for something to perfectly deserve the name ‘value’. There are no perfect deservers of the name to be had. But there are plenty of imperfect deservers of the name, and my original version is meant to capture what it takes to be one of the best of them. (But I do not say mine is the only version that can claim to do so. Doubtless there are more dimensions of semantic variation and indeterminacy than just our degree of tolerance for imperfection.) Strictly speaking, nothing shall get the name without deserving it perfectly. Strictly speaking, Mackie is right: genuine values

would have to meet an impossible condition, so it is an error to think there are any. Loosely speaking, the name may go to a claimant that deserves it imperfectly. Loosely speaking, common sense is right. There are values, lots of them, and they are what we are disposed de facto to value.

Then is my position a form of realism about values?—Irrealism about values strictly speaking, realism about values loosely speaking. The former do not exist. The latter do.

What to make of the situation is mainly a matter of temperament. You can bang the drum about how philosophy has uncovered a terrible secret: there are no values! (Shock horror: no such thing as simultaneity! Nobody ever whistled while he worked!) You can shout it from the housetops—browbeating is oppression, the truth shall make you free.26 Or you can think it better for public safety to keep quiet and hope people will go on as before. Or you can declare that there are no values, but that nevertheless it is legitimate—and not just expedient—for us to carry on with value-talk, since we can make it all go smoothly if we just give the name of value to claimants that don’t quite deserve it. This would be a sort of quasi-realism, not the same as Blackburn’s quasi-realism.27 Or you can think it an empty question whether there are values: say what you please, speak strictly or loosely. When it comes to deserving a name, there’s better and worse but who’s to say how good is good enough? Or you can think it clear that the imperfect deservers of the name are good enough, but only just, and say that although there are values we are still terribly wrong about them. Or you can calmly say that value (like simultaneity) is not quite as some of us sometimes thought. Myself, I prefer the calm and conservative responses. But so far as the analysis of value goes, they’re all much of a muchness.

26 See Hinckfuss, op. cit.