

~ 2

Intuitionism as Template: Emending Moore

NORMATIVE CONCEPTS are concepts “fraught with ought”, as Sellars put it. How do these concepts work? In this book I develop a hypothesis: normative concepts get their special characteristics, I propose, from their place in a broad kind of planning we carry out. For most of this book, I formulate this hypothesis and elaborate it, and then explore some puzzles of moral philosophy in terms of this hypothesis. Any hypothesis, though, requires testing: the one I propose will have to meet certain standards of adequacy—and so will any rival.

What these standards are is sharply contested. On one view, the young G. E. Moore set the problem of what ‘good’ means in his book *Principia Ethica* (1903). Moore’s own solution to his problem was fantastical, but an adequate theory of normative concepts, this view maintains, must pass pretty much the tests that Moore devised. On an opposing view, Moore’s tests misled generations of moral philosophers. Now, though, we can see through his tests and dismiss them as figments of semantics of a century ago. Still other philosophers continue to accept Moore’s entire program; his tests work, they think, and he drew the right, non-naturalist morals from his tests.

My own view is that Moore’s tests survive scrutiny to a remarkable degree. Not that we can accept what he says word for word—far from it. With judicious reading, extension, and revision, though, we can find in Moore the materials to construct a template that an adequate theory

of normative concepts must match. In Moore's arguments, moreover, we can find clues as to what underlies the phenomena he discovered.

Good, Moore famously insisted, is "not to be considered a natural object" (1.12, p. 14). Good and bad are *simple* objects of thought—indeed, the only simple objects of thought peculiar to ethics (1.5, p. 5). Moore acquired a philosophical following with these views, but in the decades that ensued, non-natural objects came to look spooky. Moore's arguments against naturalism, though, had a longer run; they convinced many philosophers that ethical concepts are not purely naturalistic, that we can't develop a natural science of good and bad. The great emotivists of the 1930s fully accepted Moore's claims against ethical naturalism, though they repudiated non-natural qualities. The young Moore's arguments were loose, as he later acknowledged: his "naturalistic fallacy" had proved elusive, and his "open question" test was clearly defective. The question for us now, then, is not whether Moore was right in detail, but whether he was somehow "on" to something. Do any of his arguments point to considerations we must now take seriously?

I begin, then, with a motivated reading of Moore. I adopt some parts of what Moore says and drop others, and I interpret passing remarks in light of more recent philosophical developments. My aim in part is to construct a target theory, a target that an account of ethical concepts might hit or miss. Moore, on my reading, discovered the special behavior of a class of concepts, the concepts we now call "normative". The rest of this book develops a hypothesis to account for the behaviors that Moore discovered. Moore, then, on my reading, constructs a picture with many correct features, and we can test the hypothesis of this book by seeing if it matches these features.

Eventually, I depart from Moore in important ways. Moore spoke of *good* and *bad* as the simple, non-natural objects of thought that specially figure in ethics (1.5). I myself, in most of this book, will have little to say about good and bad. I don't think these concepts are simple; rather, I follow a later non-naturalist, A. C. Ewing: the term 'good', I take it, means desirable, and desirable means something like *to be desired* or *fittingly* desired (Ewing, "A Suggested Non-Naturalist Analysis," 1939). That places the burden of explanation on the construction 'to be' in 'to be desired', or on the term 'fittingly'. The special behavior of *good* that Moore more or less uncovers I'll attribute to Ewing's concept

of what's fitting. Moore's own question, though, concerns *good*, and so for now I'll talk of good and bad.

What's at Issue?

How does Moore argue that *good* is distinct from all "natural" objects of thought? The arguments he offers are intricate, but persuasive enough that whether in detail they are correct or garbled, their force cannot have sprung from their niceties. As with many pieces of philosophical rhetoric, it is Moore's examples and the rough use he makes of them that carry the reader. Two interrelated lines of argument are the ones I find most convincing in Moore. One asks "What's at issue?" in a debate; the other appeals to coherent states of mind.

Two philosophers debate, Moore imagines; one claims that good is pleasure, the other that good is that which is desired. These are claims of identity in meaning: one philosopher claims that 'good' means the same as 'pleasant'; the other that 'good' means the same as 'desired' (or as Moore puts it, "that good just means the object of desire," p. 11).

The dispute is not verbal, Moore argues; it is not a dispute just about the English language. Whatever reasons we have, say, to go for pleasure don't rest on the meaning of a word in English—even the word 'good'. On this point Moore must be right: these two philosophers are *users* of English, who think in English or have thoughts that they express in English. The claim that good means pleasant, whatever it may amount to, would be expressed by a monolingual speaker of French as the last of the three statements below and not the first or second:

En anglais, 'good' veut dire 'pleasant',

En anglais, 'good' veut dire agréable,

Bon veut dire agréable.

If the issue, then, is not a verbal one about English, what is it? We have, I take it, a dispute about conceptual identity, conducted by us who alike use English to express our concepts.

First, then, the "What's at issue?" test. One philosopher—call him Désiré—claims that 'good' just *means* desired, and we want to test his claim. Another philosopher, Hedda, thinks that pleasure and pleasure

alone is good, whereas Désiré rejects the claim that pleasure alone is good. What's at issue? We have two assertions:

Only pleasure is good. —Hedda (\mathcal{H})

Not only pleasure is good. —Désiré ($*\mathcal{H}$)

Désiré and Hedda, Moore thinks we can see, disagree when they say these things. But Désiré can't express his disagreement with Hedda by saying:

Not only pleasure is desired. —Désiré ($*\mathcal{D}$)

For Hedda can agree with this, though she still asserts \mathcal{H} , that only pleasure is good. The first two statements contradict each other, Moore thinks we can see: $*\mathcal{H}$ contradicts \mathcal{H} . The first and the last do not: $*\mathcal{D}$ does not contradict \mathcal{H} . It follows that Désiré's two claims $*\mathcal{H}$ and $*\mathcal{D}$ don't mean the same thing: one of them contradicts what Hedda says and the other doesn't. Whether or not what's desired is always good, 'good' doesn't *mean* 'desired'. This is the argument from "What's at issue?" It asks what's at issue between Désiré and Hedda.

This argument ties in closely with a test of *conceptual coherence*. Hedda cannot both think that \mathcal{H} , only pleasure is good and that $*\mathcal{H}$, not only pleasure is good—she can't think both these things and be coherent. She can, though, coherently think \mathcal{H} and $*\mathcal{D}$: she can think that only pleasure is good, but that not only pleasure is desired. The two-person question of whether Hedda and Désiré are at odds in a set of claims boils down, then, to the one-person question: whether she could, without giving up her own claims, accept his claims and stay coherent.

Moore also made much of his "open question test", and often philosophers take this test to be crucial: whether Moore was right, they think, hinges on whether his open question test works. To test whether 'good' means desired, Moore proposed, construct the question "Is all that's desired good?" and see if the question is an open one. It's open, you'll see, whether all that's desired is good—and so 'good' and 'desired' can't mean the same. But this, critics respond, can't be a reliable test. Synonymy can be covert; if a philosopher labors to analyze a concept and dis-

cover the right analysis, the discovery won't be obvious on its face. The question whether the analysis is right will be open, because the analysis, correct though it be, is subtle.¹

The tests I'm endorsing in Moore are more demanding. To apply the "What's at issue?" argument, we need not just uncertainty, but definite findings of which claims are in disagreement and which are not—as with Hedda and Désiré. To apply the coherence test, we need the definite finding that a state of mind is coherent; it won't be enough just to find the question of its coherence open.

These tests do, though, place a great explanatory burden on two notions: (i) one person's accepting or rejecting a claim of another, and (ii) a state of mind's being conceptually coherent. Notion (i), we should note, comes also in a one-person variant over time; we have the notion (i*) a person's sticking to a claim he previously held, as opposed to rejecting it. Moore's arguments require *claims* that can be accepted or rejected at different times and by different people, and *coherence* or incoherence in accepting a set of claims. With Hedda and Désiré, after all, Moore's argument starts from a datum: that it is *coherent* to accept both \mathcal{H} and $*\mathcal{D}$, that only pleasure is good, but not only pleasure is desired—whereas it is incoherent to accept \mathcal{H} and $*\mathcal{H}$, that only pleasure is good, but not only pleasure is good. Claim $*\mathcal{D}$ is just $*\mathcal{H}$ with 'desired' substituted for 'good'. Therefore, Moore's argument concludes, the concepts *desired* and *good* are distinct. Two concepts are distinct if they offer non-equivalent possibilities of coherent acceptance or rejection. If we can't ever recognize coherence or incoherence, disagreement or compatibility, we can't apply the tests.

My claim will be that we can't live and converse without these raw

1. Notoriously, Moore claimed too that a particular fallacy, the "naturalistic fallacy", underlay many forms of naturalism. He offered many characterizations of this purported fallacy, which now strike a reader as wildly non-equivalent (see esp. secs. 1.10 and 1.12); see also Frankena, "Naturalistic Fallacy" (1939). In his talk of a "fallacy", though, Moore seemed especially concerned with this pattern of argument: Proclaim that such-and-such is "the very meaning" of the term 'good', and offer this claim negligible scrutiny. Then use this claim about the meaning of 'good' to conclude, trivially, that all and only such-and-such things are good. See Moore's discussion of Bentham (1.4). Thinking that 'good' must mean something we can state in other terms, thought Moore, closes one's mind, and makes one dismiss questions that call for careful investigation, thinking them to be quickly settled by definition.

materials of Moore's tests. A thoroughgoing, lived skepticism about meaning would paralyze thought and discourse. Without judgments of disagreement and coherence, no one could navigate a conversation. We couldn't even navigate the inner conversation of our own thoughts. Quine famously challenges whether judgments of meaning can have a clear scientific basis, and this has led to a search for ways to do philosophy without relying on notions of analyticity and conceptual identity and distinctness. What's right about Quine's conclusions and what might be wrong are complex questions, which I won't attempt to sort out. Concede to Quine, though, that what Hedda means by 'good' is empirically indeterminate. Then if you confine yourself rigorously to empirically founded judgments, you can't consider what she says; you can't agree or disagree. You can't come to reject a thought that you yourself had entertained, for you don't know which thought it was.

We respond to possible states of mind as coherent or incoherent, and this, in part, is what enables us to "track" our conversations and our own thoughts. If these intuitions exist and work systematically, then we can meaningfully ask such questions as whether *good* and *desired* figure equivalently in our conceptual intuitions. How, after all, do we navigate a discussion? How can participants and observers track it? Sometimes we don't, and listeners are reduced to bafflement. Return to Hedda and Désiré, and their initial claims:

Not only pleasure is good. —Désiré (* \mathcal{H})

Tracking their conversation requires appreciating, implicitly at least, that these two claims are in direct contradiction. Faced with Hedda's claim \mathcal{H} , Désiré could not have responded, "Yes, but not only pleasure is good." To do so would draw bafflement; this we all recognize. On the other hand, he could intelligibly have said, "Yes, but not only pleasure is desired." Hedda might agree or disagree, but part of her conversational competence is to recognize that Désiré's 'Yes, but' here is linguistically appropriate, that she hasn't, in claiming \mathcal{H} , already rejected what he is saying.

Tracking a conversation, then, requires competence with logical terms like 'not'. But this may be uncontroversial: logic is one thing, and analyticity, insofar as it outstrips logic, is quite another. The word 'not' in English is a logical term, and tracking a conversation requires hear-

ing simple logical contradictions as contradictions—with that pretty much everyone agrees. Quineans gladly accept logic as distinctive, but they deny that there are distinctively analytic truths and contradictions that are not logical truths and contradictions.

Imagine, though, this conversation.

Hedda: Only pleasure is good.

Waldo: Yes, but not only pleasure is desirable.

Hedda has every right to be baffled; Waldo's response doesn't "track". She can, of course, try to elicit what distinction he has in mind when he claims something to be "desirable" but not "good"; she can cast about for a charitable interpretation. What she cannot do is just take his words at face value, accepting or rejecting his position. Waldo isn't, in the narrowest sense, violating the logical rules of English; it isn't strictly logical terms like 'not' that cause all the problem. But he does seem to be trying to invoke a distinction that his words don't convey.

Is this phenomenon specially "conceptual"? There are many ways to draw a blank in conversation, and if Moore's diagnosis is supported by phenomena of bafflement, with signs that the conversation doesn't "track", this will require that the bafflement be distinctive in some way. It must be bafflement of a kind that is specially conceptual. Not that we need recognize it as such; Moore's taxonomies might be tenable even if we couldn't tell drawing a conceptual blank from drawing any other kind of blank. A theory of conversation and its pitfalls, after all, might classify kinds of bafflement in ways we couldn't recognize without the theory. Still, a claim that some kinds of bafflement are distinctively conceptual will need some kind of support or other.

When we draw a blank with Waldo's response, is our bafflement conceptual? What alternative hypothesis is there to invoke, what alternative that would dispense with an analytic/synthetic distinction? One is that the work is being done not by analyticity, but by manifest obviousness. A person should not be heard as accepting something that is obviously false; that is a prime maxim of interpretation. Now obviously, everything desirable is good, and so when Waldo appears to commit himself to denying this absurdity, we draw a blank. It is as if he had said, "That dog holds its breath when it chases rabbits."

Obviousness alone, though, isn't doing all the work. Clearly an important distinction can be drawn between the claim that something is desirable but not good and the claim that a dog chases rabbits without breathing. The first is false necessarily; the second is not. We can picture a dog chasing rabbits holding its breath; we know what it would *be* for a dog to do so, even if we expect it never to happen. In contrast, we don't know what it would be for a dog to sit on its own shoulders; there's no configuration, no way of coping with gravity and support, that would count as so doing. Likewise, we don't know what it would be for pleasure, or anything else, to be desirable but not good. We don't even know what it would be to *think* that something's good but not desirable.

Still, will Moorean tests work with harder cases? Philippa Foot imagines a man who insists that claspings one's hands is good, and for no reason but that it's the claspings of one's hands. Some naturalistic constraints, she concludes, are built into the very meaning of good. Does Moore have tests that let us assess Foot's claim? The hand-clasper does baffle us. (We're perhaps like the Masai boy in Kenya who flagged down a car with William Frankena in it with fellow bird-watchers: "What are they doing?" he asked the driver. "Looking for birds."—"Oh, to eat them?"—"No." Then after a pause, "Oh, they want the feathers?"—"No.") Like the Masai boy, we're baffled with the hand-clasper, because he's in a state of mind it's hard to imagine "from the inside", even in mental play-acting. Is he mixed up in his concepts, then? Not at all, I want to say—or he *may* very well not be.

But here I get ahead of my argument. Foot doesn't share my sense of the case, and so to speak with her I'd need more than my linguistic intuitions. I'd need some account of what our bafflement is, if it isn't bafflement with navigating his concepts—and the aim of this book is to develop such an account. Suppose, now, our man shows every sign of favoring hand-clasping, and no sign of having a rationale or of feeling in want of one. Then he's not mixed up in his concepts; he's got crazy views on what to do and why. Convince him that he's misusing words like 'reason', that an act's being a hand-clasping linguistically just doesn't count as a basic "reason", and he'll change his way of speaking, to be sure. But if he doesn't change his thinking on what to do, he'll still be crazy and unintelligible. And if he then does give up hand-clasping

or if he then searches for a rationale for his practice, isn't he responding to the wrong kind of consideration? Won't Foot too recognize that? (I mean here, of course, Mrs. Foot of decades ago.) If we ourselves think whether to sacrifice anything else to clasp hands, what counts in English as "good" or as a "reason" won't remotely enter in. What's wrong isn't the hand-clasper's concepts, but his grounds for acting.

The Moore-like tests I support, then, do discriminate. I argue later, for instance, that strictly naturalistic ideal observer theories don't pass them. What, though, of the kinds of analytic equivalences I do accept? *To think something good is to favor it.* This no doubt needs refinement, but let's try it even in this crude form and see if it is a valid conceptual claim. Eve tells us "I favor this action, but I don't think it's good." Or she says, "I think that this is truly the very best thing we can do, but I don't favor it." We may on further questioning manage to give some sense to her words, but her bare words don't convey what contrast she's making. If she voiced this line in speaking, we might pick up a special sense of her words from her inflection—the "good" or "best" as goody-goody, or as all too decent when indecency is called for, or some such thing. We can't, though, simply tell from her words what she means. We know the language, but she isn't quite speaking it. It's as if she said "I think it's good, but then again I don't think it's good." With luck we'll discern what she means, but we can't just read off what possibility her words allow.

Property and Concept

By some accounts of what "naturalism" is, Moore might be read almost as a modern naturalist. "*The good*", he tells us, or "that which is good", is not indefinable. By "*the good*" he means "the whole of that to which the adjective will apply, and the adjective must *always* truly apply to it" (p. 9). "I do most fully believe," he avows, "that some true proposition of the form 'Intelligence is good and intelligence alone is good' can be found" (p. 9). A true proposition of this form, he explains, would be not a definition of good, but a definition of *the good*.

What, then, does Moore mean by '*the good*'? On one apparent reading, it is the extension of the adjective 'good', the set of all and only those things that are good. For all Moore's purely conceptual argu-

ments tell us, Hedda the hedonist might be right about what this extension is. She thinks that all and only pleasant things are good, and nothing in Moore's study of concepts alone refutes her.

Moore, though, demands something stronger than this of *the* good. Suppose that Hedda is right, and suppose too that, in fact, only terrestrial beings experience pleasure. Then 'pleasure' and 'terrestrial pleasure' have the same extension: all and only terrestrial pleasant things are pleasant, and so all and only they, on Hedda's view, are good. But *the* good is pleasure, according to Hedda, not terrestrial pleasure. Being pleasant makes something good; being terrestrial doesn't. It might, after all, have been the case, though in fact it isn't, that extraterrestrial beings—beings, say, on a planet of Alpha Centauri—experienced pleasure too. This *might* have been the case, even if in fact, in the universe, there happen to be no non-terrestrial beings that are capable of pleasure, even if there never have been and never will be. *The* good, Moore tells us, is the whole of what the adjective 'good' "must *always* truly apply" to (p. 9); he thus uses the modal construction 'must *always*'. For pleasure to be *the* good, we require that all and only pleasant things are good not only as things in fact stand, but in every possible situation. In this sense, 'pleasant' and 'good' must, for pleasure to be *the* good, be coextensional necessarily.²

Is Moore, then, a believer in simple, non-natural properties? I mean not the historical Moore, the Moore of the whole of *Principia Ethica*, but a Moore we might read into the arguments and doctrines I have recounted. Moore draws from his tests a lesson about meanings: that the term 'good' means something different from any naturalistic term, from any psychological term or sociological term, for instance, from any term that can figure in purely empirical inquiries. Still, he thinks, some naturalistic formulation is coextensional with 'good' necessarily. Much work on properties in recent decades treats properties and meanings as distinct. Writers differ on how the point is best regimented, but one way is to speak of *concepts*: meanings are concepts, and concepts aren't properties. Now if Moore had spoken this way, how might he best have fit his tests and their lesson into this framework? Suppose he distinguished properties and concepts. He would then

2. Of course, this isn't a necessary truth about the terms 'good' and 'pleasant' in our community; we might have spoken differently. The point, rather, is that if Hedda is right, then necessarily, all and only pleasant things are good.

need to ask whether it is the *property* of good or the *concept* of good that his arguments and core doctrines address.

Purported examples of a property/concept distinction abound in the philosophical literature of the past few decades.³ The property of being me, I can say, is just the property of being Allan Gibbard. The two concepts I can entertain, though, are distinct: suffering amnesia but coming to know certain arcane philosophical writings, I might indignantly deny being Gibbard. Still, if I kept my logical cogency, I wouldn't deny my self-identity: "Of course I am me, but I am not Gibbard." The same might go for the concepts of being water and being H₂O, having a certain chemical structure: opponents of Lavoisier didn't deny that water is water, but they denied that water is H₂O. The concepts are distinct, at least for the naïve beginner in chemistry. As it turns out, though, the property of being water just *is* the property of being H₂O. Or so it is frequently claimed.

Properties go with necessity: in any possible situation in which I existed, I would be Allan Gibbard. I might not be called 'Allan Gibbard', for it is happenstance that, in our mouths, that name designates me; but I, as things stand, can say of any non-actual situation in which I would exist, "Allan Gibbard is who I would be." Likewise, in any possible situation water would be H₂O; if something other than H₂O behaved exactly as water behaves in everyday experience, that stuff still wouldn't be water. Identity of properties, however, though it yields necessary equivalence, does not yield *a priori* equivalence. Chemists, after all, required evidence that water is H₂O—just as I, were I amnesiac, might require evidence that I am Gibbard.

These glosses on the phenomena still no doubt require debate, but here I'll assume that they are correct and ask how to read Moore's views into such a picture. What position might his arguments and tests support? Hedda thinks that the terms 'good' and 'pleasant' are necessarily coextensive. Need she think, if she accepts Moore's arguments about meaning, that the two terms signify different properties? Nothing in the tests forces that conclusion, so long as the two terms express different concepts. Désiré, Hedda must recognize, is conceptually coherent if he denies that only pleasant things are good—just as opponents of

3. Putnam in "Meaning of Meaning" (1975) and Kripke in "Naming and Necessity" (1972) provide material for making such a distinction. Peacocke, in *A Study of Concepts* (1992), sharply distinguishes concepts from properties (p. 2).

Lavoisier were conceptually coherent in denying that all and only water is H_2O . That shows that the *concepts* of being good and of being pleasant are distinct. It doesn't show that two distinct *properties* are in play. Moore's tests show distinctness of concepts, not of properties.

If two terms stand for the same property, then they are necessarily coextensive. Does the converse hold? If two terms are necessarily coextensive, does it follow that they signify the same property? Different accounts of properties say different things on this score. Nothing I have ventured so far commits us on this issue, and so nothing so far commits my emended Moore to thinking that—on a conceptually coherent rendition of Hedda's views—'good' and 'pleasant' signify the same property. Still, I am arguing, nothing in Moore's arguments supports a conclusion that more than one property is in play. The most ontologically economical rendition of Moore's view, then, the view left after applying Occam's razor, might be that a conceptually coherent hedonist like Hedda will think that the property of being good and the property of being pleasant are one and the same.

Try reading Moore freely, then, as a non-naturalist for concepts but not for properties. This may not, of course, be what the historical Moore would have embraced if he had considered and accepted these purported distinctions. Still it fits his tests: what his tests support is claims of distinct concepts, not properties. Non-natural properties have ever since Moore seemed mysterious, and his arguments do nothing to establish them. The Moore to match in a theory of normative thought can drop non-natural properties and focus instead on normative concepts.

Concepts, we might say, can be naturalistic; these are the concepts that arise in strict empirical science and in everyday causal explanations of our experience and observations—explanations of a kind that might be elaborated into empirical science. Ethical theory, holds Moore, is not a purely empirical science; psychological and sociological claims, confined strictly to science, are not in themselves ethical claims. Claims are ethical, Moore thinks, when they involve, in an essential way, the specially ethical notions of good and bad. These notions are concepts, we can say, and they are non-naturalistic. Moore we can emend as proclaiming not non-natural properties, but non-naturalistic concepts. All *properties* are *natural*, but some *concepts* are *non-naturalistic*.

I now stipulate some terminology. If concepts, properties, and exten-

sions are distinct, then a term like ‘pleasant’ will indicate each in a different sense. The term ‘pleasant’, then, I’ll say

designates its *extension*, the set of pleasant things (that is, the set of actual things that are actually pleasant),
signifies the *property* of being pleasant, and
expresses the *concept* of being pleasant.

According to Hedda, the terms ‘good’ and ‘pleasant’ signify the same property—a natural property—but express distinct concepts. This emended Moore thinks Hedda wrong on what property the term ‘good’ signifies, but thinks, like her, that it signifies some natural property. This property, on his own doctrines, is refined and complex, involving trade-offs of values and the ways more elementary natural properties combine in organic wholes. It’s not, though, his arguments about meanings that are meant to establish this, and this difference in complexity isn’t one of natural versus non-natural. The terms ‘good’ and ‘pleasant’ both signify natural properties, but the property that ‘good’ signifies is vastly the more complex of the two. So thinks my emended Moore.

Synthetic, *A Priori* Necessity

Hedda thinks that necessarily—in any possible situation—all and only pleasure is good. How does she claim to know this? Her friend Reg might offer an answer. Reg, too, is a hedonist: he agrees with Hedda that, necessarily, all and only pleasure is good. He also offers an account of knowledge of good and bad; technically, it might be called an “ideal observer account with rigidification”. Good, Reg tells us, is whatever kind passes this test: every actual person would desire it to exist if that person were impartial, were normal as the actual run of people go, and had been aware, repeatedly and vividly, of all relevant facts. I’ll call such a person *ideal-normal*. This definition rigidifies, in that even as it applies to wildly different ways human desires might have run, it signifies a single kind—picked out in terms of what actual people are prone to desire. A puzzle for ideal observer theories is how they apply to possible situations where the distribution of human characteristics are far different. What of a world of sadists, who, when impartial, thrilled to the thought of anyone’s suffering and desired it? In thinking

of such a world, says Reg, the standard is how the normal run of *actual* people, in our actual world, would, if rendered ideal, react to the world of sadists. (Here the term ‘ideal’ is given a technical meaning; an *ideal* observer is impartial and, vividly and repeatedly, aware of all relevant facts. It isn’t built into the definition that an “ideal” observer isn’t a sadist; that’s a contingent, empirical matter.) If we normal people, rendered ideal in this sense, would want these sadists not to suffer, then their suffering is bad, not good. The dispositions of actual, normal people, says Reg, fix the property that amounts to being good. And that property, he agrees with Hedda, is the property of being pleasant. Necessarily, all and only pleasant things are good, and all and only unpleasant things are bad. Even in a world of sadists, suffering would be bad—though the pleasure others derive from contemplating it would be good. So says Reg.

Reg can’t coherently make these claims *a priori*. His hedonistic claims rest on a purported empirical finding, a finding about the dispositions of actual people. The bulk of actual people, he claims, are disposed, when rendered ideal and thinking of non-actual situations, to prefer greater net pleasure in the world to less and to have no other intrinsic preferences. This kind of thing we could only learn from experience, by investigating what actual people are like—so Reg must agree. What, then, of his claim that the better is whatever most actual people would prefer if rendered “ideal”? If he claims to establish this on the basis of experience, that will push the question of how he knows one step further. Experience of type *E*, he will be claiming, establishes that his ideal observer test works. We can ask in turn how he knows that.

Intuitionists are *foundationalists*: they claim that such a regress cannot go on forever. At some point, they say, the claim to knowledge cannot rest on experience, but must be made *a priori*. A definitional naturalist will agree, but claim further that the *a priori* claim at the end of the regress is analytic, a matter of definition. Moore the non-naturalist must reject this, for if the regress ends in an analytic truth, then the concept of good turns out to be naturalistic. Imagine Reg as a naturalist of the ilk Moore rejects: he might end the regress of empirical support at the point we have reached, and proclaim it to be secured by definition. Reg then is a definitional naturalist. The concept of being better, he says, is the concept of being preferred by any possible ideal-normal observer—where by definition, an *ideal-normal* observer is one who is normal as the run of actual people goes, except that he has been repeatedly and

vividly aware of all relevant facts. This is a naturalistic concept, and Moore thinks he can refute the view that it is the concept of good.

I haven't established that Moore is right in this. Moore may have thought he had shown that *any* version whatsoever of definitional naturalism falls before his catapults. In my own view, naturalistic analyses must be tackled case by case; I know of no argument that proves in advance that every such definition fails. I take up ideal observer definitions later.

Consider, though, an alternative kind of view that Reg might hold. He might end the regress at this point, claiming *a priori* knowledge, but not claim to have a correct analytic definition of good. What we know *a priori*, what we know but don't establish by experience, is that the better of two things is whichever would be preferred by any possible ideal-normal observer—with 'ideal-normal' defined as before. Suppose Reg takes this view of the matter. Hedda might claim to know *a priori* that the good is pleasure: that in any possible situation, all and only pleasant things would be good. If so, the grounds she cites will not include the upshot of observations; indeed she might see this as axiomatic, as clearly so on no further grounds at all. Hedda and Reg, then, are similar in important ways. Both are non-naturalists. They both think that the property of being good is the property of being pleasant. On how this is known, though, the two differ: Hedda claims to know this *a priori*, whereas Reg does not. Reg claims something else as axiomatic, *a priori* knowledge: that good things are those that would be desired by any ideal-normal observer.

Both, then, claim that the property of being good is a natural property, the property of being pleasant. Each is an intuitionist in that each rests this claim on a claim to *a priori* knowledge. The two differ, though, in the status they accord the claim that necessarily, all and only pleasant things are good. Hedda claims that we know this *a priori* and axiomatically; Reg that we know it *a posteriori*. Both claim that ethical knowledge rests on an *a priori* basis—but not on an analytic basis. If a faculty of basic, non-analytic *a priori* knowledge is called intuition, they both think that ethical knowledge depends, ultimately, on intuition.

A Template to Match

Here, then, is our Moore-like template. It is a naturalism for properties, but a non-naturalism for concepts. It denies analytic natural-

ism, the doctrine that ‘good’ can be defined, analytically, in naturalistic terms. It affirms, though, that some natural property is signified by the term ‘good’.

I don’t claim to have established this as the correct template to match. Moore did, more or less: he thought he had identified a “naturalistic fallacy” that all analytic naturalists commit, and he thought that once we surveyed a few possible versions of analytic naturalism, we could see how all such theories must fail. I myself claim that from Moore we can draw a powerful set of tools for refuting various forms of analytic naturalism. Do these tools, though, work against any form of analytic naturalism whatsoever? That I don’t claim to have established. We must scrutinize each comer and see how it fares. In later chapters, I consider a few prominent forms of analytic naturalism, analytic versions of ideal observer theories in particular. I don’t, however, discover a sweeping refutation, all in advance, of every possible form of analytic naturalism. Moreover, even refuting analytic naturalism in general, refuting it in all its possible forms, would not establish conceptual non-naturalism; it wouldn’t show that the concept of good is not a naturalistic concept. Another explanation of the phenomena, after all, might be that the concept of good is naturalistic but *sui generis*: that ethical concepts act just like other naturalistic concepts, in common sense or in the sciences, but still are not definable in terms of non-ethical concepts.⁴

I don’t, then, take it that the Moorean phenomena I have been surveying refute all forms of conceptual naturalism. Rather, in much of the rest of this book, I develop a hypothesis to explain these Moorean phenomena. My aim is to prove that a *possible* kind of thought works much as Moore maintained. To show this, I won’t consider specifically ethical concepts, and I won’t study Moore’s primitive concept of good. Instead, I take as my example the concept of being “the thing to do”. For this concept, I stipulate a built-in *to-be-domeness*, and then I study how the concept must work. It turns out to work very much as Moore, on my reading, concluded that the concept *good* works.

Indeed, I attempt something stronger: I argue that as thinkers and planners in life, we are committed to concepts that behave as Moore expects—whether or not these are concepts we have words to express.

4. Nicholas Sturgeon has pointed out this possibility to me.

For many chapters to follow, I explore how these concepts behave, these concepts to which any agent is committed. Only later do I discuss whether these concepts include the familiar normative concepts that figure in our thoughts from day to day. Elsewhere I have studied a wide range of thoughts we do have and voice: thoughts about what it *makes sense* to do, about what things are *worth* seeking and worth having in life, about what acts are *right* and what acts *wrong*, and about what acts are morally *praiseworthy* and what acts *reprehensible*.⁵ These thoughts, I have claimed, can all be explained on the pattern I explore in this book. These, however, are claims for elsewhere: My purpose in this book is to establish the possibility of such content and its intelligibility, and to show that we are all committed to such content—whether English gives us means to voice it or not.

Assume initially, then, that there are natural facts, and that we can think descriptive thoughts about them. Don't suppose, at this point, that there are any facts of what to do. Set aside for now all worries about what exactly "natural" facts are and whether we can make a sharp cleavage between them and facts that are laden with *to-be-doneness*. Seeming facts that hover near the gap will be for later. My claim now is this: that if clearly natural facts *were* all the facts there are, we would reason much as if there were facts of what to do. The concepts we use in this reasoning would behave, in many ways, like the non-naturalistic concepts proclaimed by my emended Moore. And reasoning with such concepts, as if there are such facts of what to do, is not to commit an error.

5. See my *Wise Choices* (1990), chap. 3, on what it "makes sense" to do, on right and wrong, and on the praiseworthy and reprehensible; see also my "Moral Concepts: Substance and Sentiment" (1992) and "Moral Concepts and Justified Feelings" (1993). On good and better, see my "Preference and Preferability" (1998).