6 How We Create Desire-independent Reasons for Action

I The Basic Structure of Commitment

The single most remarkable capacity of human rationality, and the single way in which it differs most from ape rationality, is the capacity to create and to act on desire-independent reasons for action. The creation of such reasons is always a matter of an agent *committing* himself in various ways. The Classical Model cannot account either for the existence or for the rational binding force of such reasons, and indeed, most of the authors in the tradition of the Classical Model deny that any such things exist. We have seen that long-term prudence is already a difficulty for the Classical Model, because on that model an agent can only act rationally on a desire that she has then and there. We saw in the case of the cigarette smoker in Denmark that it can be a requirement of rationality that an agent who lacks a desire then and there to act on her long-term prudential considerations nonetheless has a reason to do so. The Classical Model cannot account for this fact. On the Classical Model, the soldier who throws himself on a live hand-grenade in order to save the lives of his fellow soldiers is in exactly the same situation, rationally

speaking, as the child who selects chocolate over vanilla when picking a flavor of ice cream. The soldier prefers death, the child prefers chocolate. In each case, rationality is just a matter of increasing the probability of getting to a higher rung on the preference ladder.

However, I would not like such heroic cases to make it seem as if the creation of and acting on desire-independent reasons for action was somehow odd or unusual. It seems to me that we create desire-independent reasons pretty much whenever we open our mouths to talk. In this chapter we are going to examine a large class of cases where we create such reasons. It is important to state at the beginning exactly what is at issue. In some very broad sense of "want" and "desire," every intentional action is an expression or manifestation of a want or desire to perform that action. Of course, when I go to the dentist to have my tooth drilled, I do not have an urge, yen, passion, hankering after, Sehnsucht, lust, or inclination to have it drilled; but all the same, then and there, that is what I want to do. I want to have my tooth drilled. Such a desire is a motivated or secondary desire. It is motivated by my desire to have my tooth fixed. Now because every intentional action is the expression of a desire, the question arises: where do these desires come from? On the Classical Model there can be only two possibilities: either the action is one I desire to perform for its own sake or it is one I perform for the sake of some other desire I have. Either I am drinking this beer because I want to drink beer or I am drinking it to satisfy some other desire; for example, I believe it will be good for my health and I desire to improve my health. There are no other possibilities. On this account rationality is entirely a matter of satisfying desires.

It sounds a bit crass to say that every rational action is carried out to satisfy a desire, and it is therefore interesting to see the theorists in the classical tradition having so much heavy going when it comes to describing motivation. How exactly do they describe rational motivation? Bernard Williams, who thinks that there can be no external reasons and that every rational act must appeal to something in the agent's motivational set S, has this to say about the contents of S:

> I have discussed S primarily in terms of desires, and this term can be used, formally, for all elements in S. But this terminology may make one forget that S can contain such things as dispositions of evaluation, patterns of emotional reaction, personal loyalties, and various projects, as they may be abstractly called, embodying *commitments of the agent*. (My italics)[1]

A similar bifurcation is found in Davidson's characterization of "pro-attitudes." Here is what he says. "Whenever someone does something for a reason, therefore, he can be characterized as (*a*) having some sort of pro attitude toward actions of a certain kind, and (*b*) believing (or knowing, perceiving, noticing, remembering) that his action is of that kind."[2] And of his set of pro-attitudes he lists the following. It was something the agent "wanted, desired, prized, held dear, thought *dutiful*, beneficial, *obligatory*, or agreeable" (ibid., my italics). The problem with this list, as with Williams's, is that it blurs the distinction between desire-dependent and desire-independent reasons for action. It blurs the distinction between things

1. Bernard Williams, "Internal and External Reasons," in *Moral Luck*, Cambridge: Cambridge University Press, 1981, p. 105.
2. Dondald Davidson, "Actions, Reasons, and Causes," reprinted in A. White (ed.), *The Philosophy of Action*, Oxford: Oxford University Press, 1968, p. 79.

you *want* to do and things you *have* to do whether you want to or not. It is one thing to want or desire something, quite something else to regard it as "obligatory" or as a "commitment" that you have to do regardless of your desires. Why don't Williams and Davidson tell us what a commitment or an obligation is? Is it just another desire, "formally" speaking?

I think the reason that both authors appear to be struggling here is that they want to assimilate desire-independent reasons for action, which obviously exist, to desires. And the way they do this is to suggest that if we construe the set that includes desires broadly enough, then a person's commitments, obligations etc. are really members of the same set as desires. I think that blurs the crucial distinction I am trying to make between desires and desire-independent reasons for action. Why is there such a distinction? Surely, people can *want* to fulfill their obligations and keep their promises. Yes, but that is not like wanting chocolate ice cream. I want chocolate and I want to keep my promise. What's the difference? *In the case of the promise the desire is derived from the recognition of the desire-independent reason, that is, the obligation. The reason is prior to the desire and the ground of the desire. In the case of chocolate the desire is the reason.*

The points at issue in this chapter are the existence of, the nature of, the creation of, and the functioning of desire-independent reasons for action. I need to give an account of desire-independent reasons for action that meets the following conditions of adequacy:

1. The account has to be completely naturalistic. That is, it has to show how the creation and functioning of such reasons is possible for biological beasts like ourselves. We

are different from chimpanzees, but our capacities are a natural extension of other primate capacities. There must not be any appeal to anything transcendental, non-biological, noumenal, or supernatural. We are just talking about certain capacities of sweaty biological beasts like ourselves.

2. I need to specify the apparatus that enables us to create desire-independent reasons for action.

3. I need to explain how, within that apparatus, people do it, how they create such reasons. I need to state exactly the logical structure of the intentionality that underlies the creation of desire-independent reasons for action.

4. I need to explain how rationality alone makes those reasons binding on the agent. For what rational reason must the agent take into account his commitments and obligations? Why can't he just ignore them?

5. I need to explain how rational recognition of such reasons is sufficient for motivation: how such entities can rationally ground secondary desires if they are themselves desire-independent.

6. I need to explain how the apparatus and the intentionality used to answer conditions (1)–(5) is sufficient for both creation and operation of such reasons. There is no need for any help from general principles, moral rules, etc. That is, the answer to (1)–(5) must explain how desire-independent reasons for action are created and how they function without the assistance of substantive moral principles. The desire-independent reasons have to be, so to speak, self-sufficient.

Anyone familiar with the history of Western philosophy will think I have set myself a daunting task. I have seen

reviewers who describe this sort of enterprise as pulling a rabbit out of a hat. But I think that, in fact, if we can forget about the Classical Model and the whole tradition it embodies, the answer to our problems, though complex in detail, is rather simple in its basic structure.

It is important, however, that we give the explanation at the right level, because there are different levels at which these questions can be answered. There is the "phenomenological" level at which we describe how things seem to the agent when he is engaged in rational socially committed behavior, and there is the social or "societal" level at which we discuss the social institutions used in the creation of such desire-independent reasons for action, when we explain how such institutions are structured and what functions they play in the larger society.

I will say something about these levels later, but I want to begin by discussing the simplest and most basic level of intentionality. This is, so to speak, the atomic level that is prior to the molecular levels of phenomenology and sociology. In later sections I will put in more details about commitment, sincerity and insincerity, and the specific role of human institutions. But at the beginning, it is important to get clear about the simplest and most primitive forms of human commitments. What are the conditions of satisfaction of the intentional phenomena involved in the creation of commitments? Let us suppose that we have a speaker and a hearer who are both able to speak and understand a common language. We suppose that they are masters of the institutions of making statements, requests, promises, etc. In the simplest types of speech acts, where the speaker makes an assertion, a request, or a promise, for example, he imposes conditions of satisfaction on conditions of satisfaction. How exactly? Let us go

through the example of making an assertion with some care, and see what we find. Suppose a speaker utters a sentence, for example, "It is raining," and suppose he intends to make the assertion that it is raining. His intention-in-action is, in part, to produce the utterance, "It is raining." That utterance is one of the conditions of satisfaction of his intention. But if he is not just uttering the sentence, but actually *saying that* it is raining, if he actually *means* that it is raining, then he must intend that the utterance satisfy truth conditions, the conditions of satisfaction with downward direction of fit that it is raining. That is, his meaning intention is to impose conditions of satisfaction (i.e., truth conditions) on conditions of satisfaction (the utterance). His utterance now has a status function, it represents, truly or falsely, the state of the weather. And he is not neutral vis-à-vis truth or falsity, because his claim is a claim to truth. *That imposition of that sort of status function, of conditions of satisfaction on conditions of satisfaction, is already a commitment.* Why? Because the assertion was a free, intentional action of the speaker. He undertook to claim that it is raining and thus he is now committed to the truth of the asserted proposition. When he intentionally imposes conditions of satisfaction on conditions of satisfaction, in the manner of an assertion, he takes responsibility for those conditions being satisfied. *And that commitment is already a desire-independent reason for action.* For example, the speaker has now created a reason for accepting the logical consequences of his assertion, for not denying what he has said, for being able to provide evidence or justification for what he has said, and for speaking sincerely when he says it. All of these are the result of the constitutive rules for making assertions, and the speaker invokes those rules when he imposes

conditions of satisfaction on conditions of satisfaction. The creation of the commitments creates desire-independent reasons for action, and the commitment is already built into the structure of the speech act. In making an assertion the speaker presents a proposition with the downward direction of fit. But in so doing, he creates a commitment, which has the upward direction of fit. His assertion that it is raining will be true or false depending on whether it really is raining. But the commitment he makes will be satisfied only if the world really is the way he says it is, only if it is raining.

So far we have considered only assertions, but in fact all of the standard forms of speech acts with whole propositional contents involve the creation of desire-independent reasons for action, because the intentional imposition of conditions of satisfaction commits or obligates the speaker in various ways. Even requests and orders, though their propositional content refers to conditions imposed on the hearer rather than on the speaker, still commit the speaker in various ways. If I order you to leave the room I am committed to allowing you to leave the room and to wanting you to leave the room, for example.

What then is a commitment? The way to answer this question is to look at the logical structure of commitments. Commitments are factitive entities that meet our condition for reasons for action. A commitment has a propositional content and an upward direction of fit. Thus, if I have a commitment to go to San Jose next week, the propositional content is "that I go to San Jose next week," and the direction of fit is upward. The commitment is satisfied only if the world changes to match the content of the commitment, only if I actually go to San Jose. Without attempting to give "necessary and sufficient conditions"

one can say this: a commitment is the adoption of a course of action or policy (or other intentional content; one can, for example, be committed to beliefs and desires) where the nature of the adoption gives one a reason for pursuing the course. Thus, for example, I am committed to the practice of philosophy. And this commitment gives me a reason to pursue it even on hard days when things are not going well. Similarly one may be committed to the Catholic faith or to the Democratic Party. When Sally says that Jimmy is unwilling to "commit" she means he is unwilling to adopt a policy that will give him a reason for continuing in certain behavior and attitudes. Such reasons are desire-independent, though this is disguised from us by the fact that the sorts of commitments I have described are commitments to do things one may want to do anyhow. In this chapter we will be primarily concerned with a special form of commitment, where one creates a commitment to another person through the imposition of conditions of satisfaction on conditions of satisfaction.

Once we see the logical structure of commitments, it is easier to see how we can create a commitment in the performance of a speech act. Not all commitments are created by way of performing a speech act. For example, one may commit oneself to a policy just by adopting a firm intention to continue with that policy, but right now I am considering the class of commitments that are created publicly, normally directed to other people. We can create such a commitment for ourselves by imposing conditions of satisfaction on some other entity. It is harder to see how this works for assertives than it is for commissives, because in the case of an assertion we are imposing conditions of satisfaction with the downward direction of fit on the utterance, that is, we are making a truth claim. But

in making the truth claim we are also imposing commitments on *ourselves*. In making an assertion we take *responsibility* for truth, sincerity, and evidence. And such responsibilities, like commitments in general, have the upward direction of fit. These responsibilities are met only if the world is such that the utterance is true, the speaker is sincere, and the speaker has evidence for the assertion.

But why are such commitments, obligations, and responsibilities binding on the agent? Why can't he, rationally speaking, just ignore them? Why are they not social constructs like any others? Because the speaker stands in a special relation to his own assertions, in that he has created them as his own commitments. He has freely and intentionally bound himself by undertaking his commitments. He can be indifferent to the truth of someone else's assertion, because he has not committed himself. He cannot be indifferent to the truth of his own assertions, precisely because they are his commitments.

But how can any such an abstract, desire-independent commitment ever give rise to a secondary desire? How can it ever motivate? Well, ask yourself how evidence, proof, and even truth itself motivate someone to believe something that he does not want to believe? For example, many people did not want to believe Gödel's Theorem because it destroyed their research project. But once they recognized the validity of the proof, rationally speaking, they had no choice. To recognize the validity of the proof is already to recognize a reason for accepting it, and to recognize a reason for accepting it is already to recognize a reason for wanting to accept it. The lesson of this case, and of others that we will consider, is that desire-independent reasons motivate like any other reasons. Once you recognize something as a valid reason for act-

ing, that is, once you recognize a factitive entity, with you as subject and an upward direction of fit, you have already recognized it as a ground for wanting to do the thing you are committed to doing. My desire to speak the truth or keep my promise is derived from the fact that I recognize that I am making a statement or have made a promise, that statements and promises create commitments and obligations, and that I am required to fulfill my commitments and obligations, in the same way that my desire to have my tooth drilled is derived from my recognition that it needs to be fixed, and from my desire to take care of my health needs.

People tend to assume that the way desire-dependent reasons motivate secondary desires is unproblematic. But the way desire-dependent reasons motivate is no more and no less puzzling than the way desire-independent reasons motivate. I recognize that my desire to have my tooth fixed is a reason for having it drilled, and therefore a reason for wanting to have it drilled. I also recognize that the fact that I owe you money is a reason to pay it back, and therefore a reason for wanting to pay it back. In each case the recognition of a valid factitive entity with me as subject and the upward direction of fit is a reason for peforming an action and therefore a reason for wanting to perform the action.

The difficulty in seeing that there is nothing especially problematic about how desire-independent reasons can motivate derives in part from a tendency in our tradition to think that motivation must be a matter of causally sufficient conditions. It is a weakness of our tradition that we suppose that any account of motivation must show how the action is necessitated, how the agent *must* perform the action if he really has the right reasons. That mistake

derives from failing to recognize the gap. I might recognize my need to have my tooth drilled, just as I might recognize my obligation, and still not act on either reason. So in an account of the motivating force of desire-independent reasons for action, we are not trying to show that they cause actions by sufficient conditions. They do not. Neither do any other rational reasons for action.

An essential step in understanding motivation is to get clear about the relations between the third-person point of view and the first-person point of view. From the third-person point of view, every society has a set of institutional structures, and the members of that society are, in various ways, in the eyes of their fellow members, bound by the deontic structures within those institutional structures. They are bound as husbands, wives, citizens, taxpayers, etc. But to say that, is, so far, to say nothing about the first-person point of view. Why should I, as a conscious self, care in the least about what other people think I am bound or obligated to do? The answer is that from the first-person point of view, I, acting within those institutional structures, can voluntarily and intentionally create desire-independent reasons for myself. Institutional structures make it possible for me to do this, but—and this is the crucial point—obligations, commitments, and other motivators that I so create do not derive from the institution, but from my intentionally and voluntarily undertaking those obligations, commitments, and duties. Because of this fact, the recognition of these motivators can be rationally required of me as a conscious agent. This is obvious in the case of promises, and equally true, if less obvious, in the case of statements. Since I uttered the phrase "I promise," it is not open to me to say, "Yes I said that but I do not see why that constitutes making a

promise"; and once I have made the promise, it is not open to me to say, "Yes, I made the promise, but I do not see why that places me under an obligation." Similarly, if I said, "It is raining," it is not open to me to say, "Yes, I said that but I do not see why that constitutes making a statement," and once I have made a statement it is not open to me to say, "Yes, I made a statement, but I do not see why that is any commitment to its truth."

I have so far presented, rather swiftly, an overview of the main arguments that I will be presenting in this chapter. So far I have discussed them only at the most fundamental, atomic level. We will get to higher levels later, and I will restate in more detail the argument concerning the way desire-independent reasons can motivate actions. Let us see how the account of assertions presented so far meets our conditions of adequacy.

1. The account is completely naturalistic. Our abilities are an extension of more primitive animal and especially primate abilities. Apes have the capacity for intentionality, but they do not have the capacity for the second level of intentionality where they can impose conditions of satisfaction on conditions of satisfaction. They do not have the capacity to undertake a commitment to the truth of a proposition that it is raining by imposing conditions of satisfaction on conditions of satisfaction. Furthermore, they do not have the socially created institutions whereby we can do these things in ways that are recognizable to other members of our species, and consequently enable us to communicate these commitments to other members of our species.

2. The apparatus we use for the creation of desire-independent reasons for action is the set of constitutive

rules of speech acts and their realization in the semantic structure of actual human languages. Any language rich enough to allow the speaker to make an assertion, an order, or a promise will do the job. In real life the speaker and hearer will typically be involved in other institutional structures, such as money, property, nation-states, and marriages. The structures, both linguistic and nonlinguistic, are complex. But they are not mysterious, and I have described them in detail elsewhere.[3]

3. You create desire-independent reasons for action by imposing conditions of satisfaction on conditions of satisfaction. All such impositions are commitments, and all such commitments create desire-independent reasons for action. Where the condition of satisfaction makes reference to the speaker, as in the case of a vow or a promise, and the propositional content specifies some voluntary action by the speaker, there is an explicit creation of a desire-independent reason for action in the imposition of those conditions of satisfaction. In the case of the assertion, the commitment to action is only implicit, but it is a commitment nonetheless. Imposing conditions of satisfaction on the utterance imposes commitments on the speaker.

4. The commitments you undertake are binding on you, because they are *your* commitments. That is, because you freely and intentionally made the assertion and thus committed yourself to its truth, it is not rationally open to you

3. John R. Searle, *Speech Acts: An Essay in the Philosophy of Language*, Cambridge: Cambridge University Press, 1969; *Expression and Meaning*, Cambridge: Cambridge University Press, 1979; *Intentionality*, Cambridge: Cambridge University Press, 1983; and *The Construction of Social Reality*, New York: Basic Books, 1995.

to say that you are indifferent to its truth, or sincerity, or consistency, or evidence, or entailment. Recognitional rationality is enough. You simply have to recognize your own self-created commitments and their logical consequences.

5. The reason such reasons can motivate is that you created them as motivators. That is, you created a factitive entity with a propositional content that has the upward direction of fit, which is binding on you. By the exercise of your will in imposing conditions of satisfaction on conditions of satisfaction, you bound your will in the future vis-à-vis those conditions. This will become more obvious when we consider promises, but almost all speech acts have an element of promising. For a long time philosophers tried to treat promises as a kind of assertion. It would be more accurate to think of assertions as a kind of promise that something is the case.

6. Notice that I have stated the answer to conditions (1)–(5) without reference to any substantive external principles. Such principles as "you ought to tell the truth," "you ought not to lie," or "you ought to be consistent in your assertions" are *internal* to the notion of assertion. You do not need any external moral principle in order to have the relevant commitments. The commitment to truth is built into the structure of the intentionality of the assertion.

II Motivation and Direction of Fit

So far I have presented a bare bones account of how someone can create commitments and be motivated by them. In this section I want to add some more details to the account. Frankly, the account so far does not seem

to me very contentious, or even exciting. But I have to say that it faces enormous resistance. Why? A large part of the resistance comes from our peculiar philosophical tradition according to which any such account is impossible. According to this tradition, there must be a strict distinction between fact and value, between "is" and "ought." The tradition has produced endless numbers of books about the place of values in a world of facts and the sources of normativity in such a world. The same tradition contains an unhealthy obsession with something called "ethics" and "morality," and the authors are seldom really interested in reasons for action, and are too eager to get to their favorite subject of ethics. They regard facts as unproblematic, values as requiring explanation. But if you think about matters from the point of view of sweaty biological beasts like ourselves, normativity is pretty much everywhere. The world does indeed consist of facts that are largely independent of us, but once you start representing those facts, with either direction of fit, you already have norms, and those norms are binding on the agent. All intentionality has a normative structure. If an animal has a belief, the belief is subject to the norms of truth, rationality, and consistency. If an animal has intentions, those intentions can succeed or fail. If an animal has perceptions, those perceptions either succeed or fail in giving it accurate information about the world. And the animal cannot be indifferent to truth, success, and accuracy, because the intentional states in question are the states of that very animal. If *you* have a belief, I may be indifferent to the truth or falsity of your belief, but if *I* have a belief I cannot be similarly indifferent, because it is my belief and the normative requirement of truth is built into the belief. From the point of view of the animal, there is no escape

from normativity. The bare representation of an *is* gives the animal an *ought.*

What is special about human animals is not normativity, but rather the human ability to create, through the use of language, a *public* set of commitments. Humans typically do this by performing public speech acts where the speaker intentionally imposes conditions of satisfaction on conditions of satisfaction. These speech acts are made possible by the existence of institutional structures that the speaker uses to perform meaningful speech acts and to communicate them to other speakers/hearers. Using this apparatus the speaker can undertake commitments when he imposes conditions of satisfaction on conditions of satisfaction. Indeed there is no way to avoid undertaking commitments. The speech act of asserting is a commitment to truth, the speech act of promising is a commitment to a future action. Both arise from the fact that the speaker imposes conditions of satisfaction on conditions of satisfaction. Speech acts commit the speaker to the second set of conditions of satisfaction. In the case of an assertion, he is committed to the truth of the assertion, in the case of a promise, he is committed to carrying out the act that he has promised to perform.

Once a motivation is created, its recognition provides an internal reason for action. It is important to get clear about this point. The acceptance of any external motivator, however crazy, can provide an agent with an internal reason for an action. If I irrationally become convinced that there is a tiger hiding behind my desk, then I have accepted the existence of a danger, and I consequently have a reason for acting, however irrational my reason may be. The point, however, about the desire-independent reasons for action is that their acceptance is rationally

required as a matter of recognitional rationality, once the agent has intentionally and freely created the reason in question.

Consider the case I discussed earlier where I make a statement that it is raining. Whenever I make a statement I have a reason to speak truthfully. Why? Because a statement simply *is* a commitment to the truth of the expressed proposition. There is no gap at all between making a statement and committing oneself to its truth. That is, there are not two independent features of the speech act, first the making of a statement and second committing myself to its truth; there is only making the statement, which is *eo ipso* a commitment to truth. Suppose you ask me, "What's the weather like outside?" And I say "It's raining." I have *thereby* committed myself to the truth of the proposition that it is raining. My commitment to truth is most obvious in cases where I am lying. If I don't in fact believe that it is raining, but I lie and say, "It's raining," my utterance is intelligible to me as a lie precisely because I understand that the utterance commits me to the truth of a proposition I do not believe to be true. And the lie can succeed as a lie precisely because you take me to be making a statement and therefore committing myself to the truth of the expressed proposition. A similar point can be made about mistakes. Suppose I am not lying but am genuinely mistaken. I sincerely said it is raining, but all the same it is not raining. In such a case there still is something wrong with my speech act, namely, it is false. But why is that wrong? After all, for every true proposition there is a false one. It is wrong because the aim of a statement is to be true, and this one fails, because it is false. When I make a statement I commit myself to

its truth, and here my mistake makes me fail in the commitment.

There is no way that the Classical Model can account for these simple facts. The Classical Model is forced to say that there are two separate phenomena, the institution of statement making and then, external to that, the principle that one should try to speak the truth. What reason have I to try to tell the truth when making a statement? The classical theorist is forced to say that I have *no reason at all just in virtue of making a statement*. The only reason I could have would be that I felt there would be bad consequences if I lied, or that I hold a moral principle, which is logically independent of making a statement, to the effect that falsehood is wrong, or that I just felt an inclination to tell the truth, or had some other reason external to making the statement. On the Classical Model all such reasons are independent of the nature of statement making as such. I am claiming, on the contrary, that there is no way to explain what a statement is without explaining that the commitment to truth is *internal* to statement making.

But why is the commitment to truth internal to statement making? Why couldn't we have a different sort of institution of statement making, where we make statements, but are not committed to their truth? What is the big deal about commitment? Well, in a sense you can perform speech acts without their normal commitments. That is what happens in works of fiction. In works of fiction nobody holds the author responsible for the truth of the utterances that she makes in the text. We understand those cases as derivative from, and parasitic on, the more fundamental forms, where the commitments are to the truth conditions of the actual utterance. So, to repeat the

question, why? And the answer follows from the nature of meaning itself. The reason why I am committed to the truth of the claim that it is raining when I say that it is raining is that, in making the utterance that it is raining, I have intentionally imposed certain conditions of satisfaction on that utterance. Assuming I am not just practicing my pronunciation, or rehearsing for a play, or reciting a poem, when I seriously assert that it is raining, I am committed to the truth of the proposition, because I have intentionally imposed the commitment to that truth on the utterance when I intentionally imposed the conditions of satisfaction that it be raining on the conditions of satisfaction of my intention-in-action that that intention-in-action should produce the sounds, "It is raining." And, to repeat, what makes it possible for me to do that in a publicly accessible manner is the fact that I am a participant in the human institution of language and speech acts.

Now I want to apply some of these lessons to practical reason as it is more traditionally construed. In many cases of practical reason, one creates a reason now for performing an act in the future. I believe the only way to understand how voluntary rational action can create reasons for future actions is to look at the matter from close up. So, let us consider the sorts of cases that happen in everyday life. Suppose I go into a bar and order a beer. Suppose I drink the beer and the time comes to pay for the beer. Now the question is, granted the sheer fact that I intended my behavior to place me under an obligation to pay for the beer, must I also have a reason independent of this fact, such as a desire to pay for the beer, or some other appropriate element of my motivational set, in order to have a reason to pay for the beer? That is, in order to know if I have a reason to pay for the beer, do I first have to scruti-

nize my motivational set to see if there is any desire to pay for the beer, or to see if I hold any general principles about paying for beer that I have drunk? It seems to me the answer is, I do not. In such a case, by ordering the beer and drinking it when brought, I have already intentionally created a commitment or obligation to pay for it, and such commitments and obligations are species of reasons.

It is an absurdity of the Classical Model that it cannot account for such an obvious case. As in the case of truth telling, the defender of the Classical Model is forced to say that I have a reason to pay for the beer only if I can locate the relevant desire in my "motivational set." In opposition to this I want to claim that in this situation I have simply created a reason for myself to pay for the beer by ordering the beer and drinking it.

What exactly are the formal features of the situation that have enabled me to create such a reason? What exactly are the truth conditions of the claim: Agent A has a desire-independent reason to perform act *X* in the future? What fact about him makes it the case that he has such a reason? Well, one sort of fact that would be sufficient is: Agent A has *created* a desire-independent reason for himself to perform act *X* in the future. So our question now boils down to: how does one go about such a creation? I have already answered that question as a logical question about conditions of satisfaction, but let us now consider it "phenomenologically." How did it seem to Agent A when he ordered the beer? Well, if I am the agent, the way that it seems to me is this: I am now performing an act such that I am in that very act trying to get the man to bring me a beer on the understanding that I am under an obligation to pay for it if he brings it. But if that is the intention, then, by this very performance, if the man brings the beer, I

have made it the case that I now have an obligation, and therefore a reason, which will be a reason for me to act in the future, and that reason that I now create will be independent of my other future desires. In such a case, a sufficient condition for an act to create a reason for me is that I intend that it create a reason for me.

The formal mechanism by which I created the obligation is exactly parallel to the formal mechanism by which I created a commitment in the case of statement making. In this case, however, I imposed conditions of satisfaction on my utterance, which had an upward direction of fit. I undertook an obligation to do something. It is hard to see this, because I did not do this explicitly in the utterance. I just said, "Bring me a beer," and that utterance has the conditions of satisfaction with the upward direction of fit, that the hearer should bring me a beer. But the total understanding of the situation, which we will have occasion to explore in detail when we consider promising, is that I have also imposed conditions of satisfaction on myself, on my future behavior. And I have imposed these in the form of a conditional obligation. Obligations have the upward, or world-to-obligation, direction of fit. The obligation is satisfied or fulfilled only if the world changes, typically in the form of the behavior of the person who has the obligation, to match the content of the obligation. Obligations, therefore, are a species of external motivators. Typically their existence is epistemically objective, though because they are always created by human beings, and exist only relative to the attitudes of human beings, they are ontologically subjective. And as we have had occasion to see over and over, ontological subjectivity does not imply epistemic subjectivity. It can be a plain matter of fact that I am under an obligation, even though

the creation and the existence of the obligation are observer-relative.

The presupposition of the freedom of the agent is crucial to the case as I have described it. From the first-person point of view, by freely undertaking to create a reason for myself, I have already manifested a desire that such and such be a reason for me. I have already bound my will in the future through the free exercise of my will in the present. In the end all these questions must have trivial answers. Why is it a reason? Because I created it as a reason. Why is it a reason *for me*? Because I have freely created it as a reason for me.

In the discussion of the gap in chapters 1 and 3, we found that all effective reasons are agent created. But the peculiarity of the creation of desire-independent reasons for future actions is that I now, through the exercise of an effective reason, have created a potentially effective reason for me to act in the future. The philosophical tradition has the problem exactly back to front. The problem is not, "How could there be desire-independent reasons for me?"; the problem is rather, "How could anything be a reason of any kind for me that I did not create as a reason for me, including desire-independent reasons?" In the performance of a voluntary action, there is a gap between the causes and the actual carrying out of the action, and that gap is crossed when I simply perform the action; and in this case, the performance of the action is itself the creation of a reason for a subsequent action.

As far as motivation is concerned, in the cases I have described *the reason can be the ground of the desire and not conversely*. In ordinary English the correct description of this case is, "I want to pay for it because I have an obligation to pay for it." And the connection between reason,

rationality, and desire is as follows: the recognition of something as a binding obligation is already the recognition of something whose ontology is that of an external motivator, that is, an entity that has the upward direction-of-fit. To recognize the validity of such an entity is already to recognize a reason for acting. And the recognition of something as a reason for acting is already the recognition of that thing as a reason for desiring to perform the action.

III Kant's Solution to the Problem of Motivation

Kant, in *Groundwork of the Metaphysic of Morals*,[4] faced a problem that is formally similar to the one I am discussing. My problem is, how can desire-independent reasons actually motivate actions, if every action is the expression of a desire to perform that action? Kant phrases his problem in the form, "How can pure reason be practical?" And he explains that by saying that it is the question of why we can take an *interest* in the Categorical Imperative. An interest is that in virtue of which reason becomes practical, that is, it becomes a cause determining the will to action. It seems to me that Kant's answer to this question is inadequate. Here is what he says: "If we are to will actions for which reason by itself prescribes an 'ought' to a rational, yet sensuously effected, being, it is admittedly necessary that reason should have a power of *infusing a feeling of pleasure* or satisfaction in the fulfilment of duty, and consequently that it should possess a kind of causality by which it can determine sensibility in accordance with rational principles" (p. 128). So, on Kant's view, pure rea-

4. Immanuel Kant, *Groundwork of the Metaphysic of Morals*, New York: Harper Torchbooks, 1964.

son has to cause a feeling of pleasure, and it is only because of that feeling of pleasure that we are actually able to act in accordance with the dictates of pure reason. Kant admits that it is totally unintelligible to us how pure reason could ever cause such a feeling of pleasure, because we can only discover cause and effect relations among objects of experience, and pure reason is not an object of experience.

I think this is a bad argument. Kant's claim is that we could not act on a desire-independent reason for action unless, somehow or other, we would get a "feeling of pleasure" from doing so. I think Kant fails to understand direction of fit. That is, I think we can perform many actions in which there is no "feeling of pleasure," only the recognition that we have a valid reason for doing them. I no more have to have a "feeling of pleasure" when I get my tooth drilled than I have to have a feeling of pleasure when I keep my promises. I might get some satisfaction out of the tooth drilling and from the fulfillment of my promise, but it is not logically necessary that I get any such feeling in order for me to have my tooth drilled or to keep my promise. *On the view that I am presenting, the recognition of the validity of the reason is enough to motivate the action.* You do not need to have any extra pleasure, desire, or satisfaction. The motivation for performing the action is precisely the motivation for wanting to perform the action.

This is an absolutely crucial point, both for Kant's argument, as well as for the argument of this book, and indeed for the debate about the Classical Model in general. Kant, though he attacks the Classical Model in various ways, accepts one of its worst features. Kant assumes that I could not intentionally and voluntarily perform an

action here and now, unless I got a "feeling of pleasure" here and now, in the performance of that action. If every action is really done to satisfy a desire, and if every action is itself the expression of a desire to perform that action, then there must be some desire satisfaction in the performance of any action. But this is a nest of confusions, and I intend now to sort them out. First let us consider the cases where an action is done to satisfy a desire. I get my tooth drilled to satisfy my desire to have it fixed. And I get it drilled because I want then and there to get it drilled. But it does not follow that there need be any "feeling of pleasure" in any sense at all in my intentional action. The primary desire to get my tooth fixed can motivate a secondary desire to have it drilled, which in turn can motivate the action. But the pleasure or satisfaction that I get from having a repaired tooth does not carry over to the activity of getting it drilled, nor need it. This is a case where I have a desire-dependent reason for desiring something, but the way that the desire-dependent reason grounds the secondary desire is exactly the same way that a desire-independent reason grounds a secondary desire. My desire to keep my promise derives from the desire-independent fact that I have made a promise, and therefore have an obligation. But it is no more necessary that I derive a feeling of pleasure from keeping my promise in order that I intentionally perform the action of keeping my promise, than it is necessary that I derive a feeling of pleasure from having my tooth drilled in order that I satisfy my primary desire of getting my tooth fixed. Kant's mistake makes fully explicit a mistake that is only implicit in most of the authors in the Classical tradition. If every action is the expression of a desire to perform that action, and every successful action results in the satisfaction of

desire, then it seems that the only thing that can motivate an action is desire satisfaction, that is, a feeling of pleasure. But this is a fallacy. From the fact that every action is indeed the expression of a desire to perform that action, it does not follow that every action is done for the *purpose* of satisfying a desire, nor does it follow that actions can be motivated only by desire satisfaction, in the sense of a feeling of pleasure.

IV Promising as a Special Case

Discussions of these issues usually spend a lot of time on promising, but I am trying to emphasize here that the phenomenon of agent-created desire-independent reasons is pervasive. You could not begin to understand social life without it, and promising is only a special and pure kind of case. However, the history of the debates about promising is revealing, and I will be able to explain better what I am arguing for if I explain the obligation to keep a promise and expose some of the standard mistakes. The question is: what reason do we have for keeping a promise? And to that the obvious answer is: promises are by definition creations of obligations; and obligations are by definition reasons for action. There is a follow-up question: what is the *source* of the obligation to keep a promise?

There is no way that the Classical Model can account for the fact that the obligation to keep a promise is internal to the act of promising, just as the commitment to truth telling is internal to the act of statement making. That is, promising is by definition undertaking an obligation to do something. The tradition is forced to deny this fact, but in order to deny it, the defenders of the Classical Model are typically forced to say some strange, and I believe

mistaken, things. In this section I offer a brief list of the most common mistakes I have encountered.

There are three common but I believe mistaken claims that can be disposed of quickly. The first is to suppose there is some special *moral* obligation to keep a promise. On the contrary, if you think about it you will see that there is no special connection between promising and morality, strictly construed. If I promise to come to your party, for example, that is a social obligation. Whether it is a moral obligation as well would depend on the nature of the case, but for most parties I go to it would not be a moral obligation. Often we make promises where some grave moral issue is concerned but there is nothing about promising as such that entails that any promise at all involves moral issues. There is nothing in the practice of promising as such that guarantees that every obligation to keep a promise will be grave enough to be considered a moral obligation. One may make promises over matters that are morally trivial.

A second, related mistake is to suppose that if you promise to do something evil there is no obligation at all to keep the promise. But this is obviously wrong. The correct way to describe such cases is to say that you do indeed have an obligation to keep the promise but it is overridden by the evil nature of the promised act. This point can be proved by the method of agreement and difference: there is a difference between the person who has promised to do the act and the person who has not. The person who has made the promise has a reason that the person who has not made the promise does not have.[5]

5. In law, a contract to do something illegal is considered null and void and cannot be enforced in court. That is not because there was no contract, but because the law voids it.

A third, and I believe the worst of the three mistakes, is to suppose that the obligation to keep a promise is only a prima facie obligation, as opposed to a flat outright obligation. This view was formulated (by Sir David Ross)[6] to try to get around the fact that obligations typically conflict and you often can't fulfill them all. When obligation A overrides obligation B, says Ross, B is only a prima facie obligation, not an honest-to-john outright obligation. I have argued in detail elsewhere[7] that this view is confused, and I won't repeat the arguments here except to say that when B is overridden by some more important obligation, this does not show that B was not an all-out, unconditional, etc. obligation. You can't override it if there there is nothing really there to override in the first place. "Prima facie" is an epistemic sentence modifier, not a predicate of obligation types, and could not possibly be an appropriate term for describing the phenomenon of conflicting obligations, where one is overridden by another. The theory of "prima facie obligations" is worse than bad philosophy, it is bad grammar.

I believe the following are the most common serious mistakes about the obligation to keep a promise, and they all derive in their different ways from an acceptance of the Classical Model:

Mistake number 1: The obligation to keep a promise is prudential. The reason for keeping a promise is that if I don't I will not be trusted in the future when I make promises.

---

6. W. D. Ross, *The Right and the Good,* Oxford: Oxford University Press, 1930, p. 28.
7. John R. Searle, "Prima Facie Obligations," in Zak van Straaten (ed.), *Philosophical Subjects: Essays Presented to P. F. Strawson,* Oxford: Oxford University Press, 1980, pp. 238–260.

Famously, Hume held this view. But it is subject to a decisive, and equally famous, objection: on this account, in cases where no living person knows of my promise, I would be under no obligation at all to keep it. On this view the deathbed promise, made by the son in private to his dying father, would involve no obligation at all because the son need not tell anyone about the promise.

Furthermore, why would I not be trusted in the future? Only because I undertook an obligation and failed to carry it out. The failure to fulfill obligations as a ground for mistrust is quite unlike the mere fact of disappointing expectations. For example, Kant famously took his walks so regularly that his neighbors could set their clocks by him. But if he failed to take his walk at the regular time, he may have disappointed, but he would not have inspired mistrust in the way that a person who reneges on his obligations would. In the promising case the mistrust arises not just from the failure of an expectation, but from the fact that the promissor gave his word.

Mistake number 2: The obligation to keep a promise derives from the acceptance of a moral principle to the effect that one ought to keep one's promises. Without such an acceptance the agent has no reasons, except perhaps prudential reasons to keep a promise.

The mistake here is the same as the mistake we found in the case of the commitment to truth when making a statement. The Classical Model tries to make the obligation in promising external to the act of promising, but then it becomes impossible to explain what a promise is, just as it becomes impossible to explain what a statement is if one tries to make the relation between stating and committing oneself to the statement's truth purely external. That

is, the decisive answer to this objection is to point out that the relations between promising and obligations are internal. By definition a promise is an act of undertaking an obligation. It is impossible to explain what a promise is except in terms of undertaking an obligation.

Just as we saw in the case of statement making, that the commitment to truth is most obviously revealed in the case of the person who deliberately lies, so in the case of promising we can show that the obligation is internal to the act of promising most obviously in the case of the person who makes an insincere promise. Suppose I make an insincere promise, a promise I have no intention to keep. In such a case my act of deception is fully intelligible to me, and may later be seen by the promisee as a dishonest act, precisely because it is understood that when I made the promise I was binding myself, undertaking an obligation, to do the thing I promised to do. When I make a promise I am not hazarding a guess or making a prediction about what is going to happen in the future; rather I am binding my will as to what I am going to do in the future. *My dishonest promise is intelligible to me as a promise in which I undertook an obligation without any intention to fulfill the obligation I have undertaken.*

Mistake number 3 (this is a more sophisticated variant of number 2): If obligations really were internal to promising then the obligation to keep a promise would have to derive from the institution of promising. The fact that someone made a promise is an institutional fact, and any obligation would have to derive from the institution. But then what is to prevent any institution from having the same status? Slavery is as much an institution as promising. So if the view that promises create desire-independent

reasons were right, then the slave would have as much an obligation as does the promissor, which is absurd. That is, the desire-independent view of promising leads to absurd results and so must be false. The correct way to see matters is to see that the institution is indeed the ground of the obligation but only because independently of the institution we accept the principle that one ought to keep one's promises. Unless you approve of the institution or somehow endorse it or favorably evaluate it, there would be no obligation of promising. We are typically brought up to keep our promises and thus to adopt a favorable attitude toward the institution, so we fail to notice that our endorsement of the institution is the essential source of the obligation. As institutions, promising and slavery are on all fours; the only difference as far as our present debate is concerned is that we happen to think the one is good, the other bad. But the obligation is not internal to the act of promising, it derives externally from the attitude that we have toward the act of promising. The only way the obligation of promising could be created is that we accept the principle "Thou shalt not break thy promise."

This objection encapsulates the view of the Classical Model on this issue. The simplest answer to it is this: *The obligation to keep a promise does not derive from the institution of promising*. When I make a promise, the institution of promising is just the *vehicle*, the tool that I use to create a reason. The obligation to keep a promise derives from the fact that in promising I freely and voluntarily create a reason for myself. The free exercise of the will can bind the will, and that is a logical point that has nothing to do with "institutions" or "moral attitudes" or "evaluative

utterances." This is why the slave does not have any reason to obey the slave owner, except prudential reasons. He has not bound his will by an exercise of his freedom. Viewed externally, the slave may look exactly like the contract laborer. They might even be given the same rewards. But internally it is quite different. The contract laborer has created a reason for himself that the slave has not created. To think that the obligation of promising derives from the institution of promising is as mistaken as to think that the obligations I undertake when I speak English must derive from the institution of English: unless I think English is somehow a good thing, I am under no obligations when I speak it. On the Classical Model, the obligation to keep a promise is always something external to the promise itself. If I have an obligation to keep a promise it can only be because I think (a) that the institution of promising is a good thing, or (b) I hold a moral principle to the effect that one ought to keep one's promises. There is a simple refutation of both of these views: they have the consequence that in the absence of either of these conditions, there would be no obligation whatever to keep a promise. So, for someone who did not think the institution of promising was a good thing, or for someone who did not hold a moral principle that one ought to keep one's promises, there is no reason whatever to keep a promise. I believe that is absurd, and I have been pointing out its absurdity at various points throughout this book.

Mistake number 4: There are really two senses of all these words, "promise," "obligation," etc., a descriptive and an evaluative sense. In the descriptive sense, when we use these words, we are just reporting facts and not actually endorsing any reasons for action. When we use them in

an evaluative sense, more is involved than just stating facts, for in these cases we must make some moral judgement, and such moral judgements can never follow from the facts by themselves. So, really, there is a systematic ambiguity in the whole discussion. The ambiguity is between the descriptive and the evaluative meanings of the words.

I will be brief in answering mistake number 4. There are no such two senses of these words any more than there are two senses of "dog," "cat," "house," or "tree." Of course one can always use words in a way that does not involve the normal commitments. Instead of saying "That's a house," I can say "That's what they call 'a house,'" in which case, I don't commit myself one way or another to whether it is actually a house (though I do commit myself to some people calling it that). Now, similarly, if I say "He made a promise" or "He undertook an obligation," I can put quotation marks around the words "promise" and "obligation" and thus remove the commitment carried by the literal meaning of the words. But this possibility doesn't show that there are two senses to any of these words or that there is some ambiguity in their literal use. The literal meaning of "promise" is such that someone who has made a promise has thereby undertaken an obligation to do something. It is an evasion of these matters to try to postulate extra senses of these words.

V Generalizing the Account: The Social Role of Desire-independent Reasons

So far in this chapter I have tried to describe what I call the atomic structure of the creation of desire-independent

reasons for action, and I have discussed some of the special features of assertions and promises with emphasis on criticizing the philosophical tradition in its discussion of the institution of promising. I have also briefly discussed the "phenomenological level" of desire-independent reasons for action, where one acts on the understanding that one's action will create a reason for oneself to do something in the future. I now want to try to state a more general account of the role of desire-independent reasons in social life in general, at a higher level than the level of the atomic structure. I want, among other things, to explain why the creation of desire-independent reasons by free, rational selves in possession of a language and operating within institutional structures is pervasive. This is what happens when you get married, order a beer in a bar, buy a house, enroll in a college course, or make an appointment with your dentist. In such cases you invoke an institutional structure in such a way that you create a reason for yourself to do something in the future regardless of whether in the future you have a desire to do that thing. And in such cases it is a reason for you because you have voluntarily created it as a reason for you.

A general account of the role of reasons in practical rationality involves understanding at least the following five features: *(1) freedom; (2) temporality; (3) the self, and with it the first-person point of view; (4) language and other institutional structures; and (5) rationality.* Let us consider each in order.

Freedom

I have already argued that rationality and the presupposition of freedom are coextensive. They are not the same

thing, but actions are rationally assessable if and only if the actions are free. The reason for the connection is this: *rationality must be able to make a difference.* Rationality is possible only where there is a genuine choice between various rational and irrational courses of action. If the act is completely determined then rationality can make no difference. It doesn't even come into play. The person whose act is entirely caused by beliefs and desires, à la the Classical Model, is acting compulsively outside the scope of rationality altogether. But the person who freely *acts on* those same beliefs and desires, who freely makes them into *effective* reasons, acts within the realm of rationality. Freedom of action, the gap, and the applicability of rationality are coextensive.

Acting freely, I can, by imposing conditions of satisfaction on conditions of satisfaction, create a reason that will be a reason for me to do something in the future, regardless of whether I feel like doing it when the time comes. The ability to bind the will now can create a reason for the future act only because it is a manifestation of freedom.

Temporality

Theoretical reason statements are untensed in a way that practical reason statements are inherently tensed. "I am going to do act A because I want to make it the case that B" is essentially future referring, in the way that "Hypothesis H is substantiated by evidence E" is not essentially tensed at all. It is timeless, although of course in particular instances, it may make reference to particular historical situations. For nonhuman animals, there really are only immediate reasons, because without language you cannot order time.

The self and the first-person point of view

In the cases we will be considering, it is essential to see that we are examining the logical structure of the behavior of rational selves engaged in creating reasons for themselves. No external or third-person point of view can explain the processes by which a free agent can create a reason now that will be binding on him in the future, regardless of how he may feel in the future.

Language and other institutional structures

In order to create desire-independent reasons an agent has to have a language. One can imagine primitive pre-linguistic beings imposing conditions of satisfaction on conditions of satisfaction. But the systematic creation of such reasons, and their communication to other people, requires conventional symbolic devices of the sort that are characteristic of human languages. Furthermore, social relations require that we be able to represent the deontic relations involved in the creation of desire-independent reasons for action, and we also need language to order time in the required way. That is, we have to have ways of representing the fact that one's present action creates a reason for a future action, and we have to have linguistic ways of representing the temporal and deontic relations in question.

In addition to language narrowly construed, that is, in addition to such speech acts as statement making or promising, there are extralinguistic institutional structures that also function in the creation of desire-independent reasons. So, for example, only if a society has the institution of property can there be desire-independent reasons

involving property, and only if a society has the institution of marriage can there be desire-independent reasons involving the institution of marriage. The point, however, which must be emphasized over and over, is that the reason does not derive from the institution, rather the institution provides the framework, the structure, within which one creates the reason. The reason derives from the fact that the agent binds her will through a free and voluntary act.

Rationality

In order that the practice of creating desire-independent reasons can ever be socially effective, it must be effective in virtue of the rationality of the agents involved. It is only because I am a rational agent that I can recognize that my previous behavior has created reasons for my present behavior.

Combining all five elements

How let us try to put these points together into a general account. To begin with, how can we organize time? The obvious answer is that we do things now that will make things happen in the future in a way they would not have happened if we did not act now. That is why we set our alarm clocks. We know we have a reason to get up at 6:00 A.M., but we also know that at 6:00 A.M. we will not be able to act on that reason because we will be asleep. So by setting the alarm clock now, we will make it possible to act on a reason in the future. But suppose I don't have an alarm clock and I have to try to get some other person to wake me up. What is the difference between setting an

alarm clock for 6:00 A.M. and asking someone to wake me up at 6:00 A.M., for example? In both cases I do something now to make it the case that I will wake up at 6:00 A.M. tomorrow. The difference is that in the alarm clock case only causes are created, whereas in the latter case, new reasons for action are created. How? Well, there are different sorts of cases. If I don't trust the person in question I might say, "If you wake me up at 6:00 A.M. I will give you five dollars." In that case I have made a promise, a conditional promise to give the other person five dollars, and, if he accepts the offer, he has promised to wake me up on the condition that I pay him five dollars. This is typical of contracts. Each party makes a conditional promise, conditional on receiving a benefit from the other party.

In the more realistic case I simply extract from him a promise to wake me up. I say, "Please wake me up at 6:00 A.M.," and he says, "OK." In that context he has made an unconditional promise and created a desire-independent reason.

In a third sort of case, no promise need be made at all. Suppose I do not trust the person at all, but I know that he makes his breakfast everyday at 6:00 A.M. I simply position all the breakfast food so that he can't get at it without waking me up. I take it in my room and lock the door, for example. To get breakfast he has to bang on my door to wake me up. Now this third sort of case also creates a reason to wake me up, but this one is a prudential or desire-dependent reason. He has to reason: "I want breakfast, I can't have breakfast unless I wake him up, so I will wake him up."

All three of these methods might on occasion work equally well, but I want to call attention to what a bizarre

case the third one is. If the only way we could get cooperation from other people was by getting them in a position where they, independently of us, want to do what we want them to do, most forms of human social life would be impossible. *In order that we can organize time on a social basis it is necessary that we create mechanisms to justify reasonable expectations about the future behavior of members of the community, ourselves included.* If we only had desires, in the manner of Köhler's apes, we would never be able to organize time in a way that would enable us to organize our own behavior, and to coordinate with other selves. In order to organize and coordinate our behavior, we need to create a class of entities that will have the same logical structure as desires, but will be desire-independent. We need, in short, to create a class of external motivators that will provide a reason for an action—that is, a propositional content with an upward direction-of-fit, and the agent as subject. The only way that such entities can be binding on rational selves is precisely if the rational selves freely create them as binding on themselves.

Let us turn to the role of language and other institutional structures. There are many features of institutional facts that require analysis; I have elsewhere tried to give an analysis of several of them and I won't repeat it here.[8] However, there is one feature that is essential for the present discussion. In the case of institutional facts, the normal relationship between intentionality and ontology is reversed. In the normal case, what *is* the case is logically prior to what *seems to be* the case. So, we understand that the object seems to be heavy, because we understand what it is for an object to be heavy. But in the case of institu-

8. *The Construction of Social Reality*, New York: The Free Press, 1995.

tional reality, the ontology derives from the intentionality. In order for a type of thing to be money, people have to think that it is money. But if enough of them think it is money and have other appropriate attitudes, and act appropriately, and if the type of thing satisfies all the other conditions set by their attitudes, such as not being counterfeit, then it is money. If we all think that a certain sort of thing is money and we cooperate in using it, regarding it, treating it as money, then it is money. In this case, "seems" is prior to "is." I cannot exaggerate the importance of this phenomenon. The noises coming out of my mouth, seen as part of physics, are rather trivial acoustic blasts. But they have remarkable features. Namely, we think they are sentences of English and that their utterances are speech acts. If we all think of them as sentences and speech acts, and if we all cooperate in using, interpreting, regarding, responding to, and generally treating them as sentences and speech acts, then they are what we use, regard, treat, and interpret them as. (I am being very brief here. I do not wish to suggest that these phenomena are in any way simple.) In such cases we create an institutional reality by treating a brute reality as having a certain status. The entities in question—money, property, government, marriage, universities, and speech acts—all have a level of description where they are brute physical phenomena like mountains and snowdrifts. But by collective intentionality we impose statuses on them, and with those statuses we impose functions that they could not perform without that imposition.

The next step is to see that in the creation of these institutional phenomena we can also create reasons for action. I have a reason for preserving and maintaining the rather uninteresting bits of paper in my wallet, because I know

that they are more than just bits of paper. They are valuable pieces of United States currency. That is, given the institutional structure, there are whole sets of reasons for actions that could not exist without the institutional structure. So, "it seems to be the case" can create a set of reasons for action, because what seems to be the case (appropriately understood) is the case, where institutional reality is concerned. If I borrow money from somebody, or order a beer in a bar, or get married, or join a club, I use institutional structures to create reasons for action and the reasons exist within institutional structures.

But so far this doesn't answer our crucial question, namely, how can we use such structures to create desire-independent reasons? I have very good reasons for wanting money, but they are all desire-dependent, because they derive from the desires I have for the things I can buy with the money. But what about the obligations I have to pay money? Or pay my debts to other people? Or fulfill my promises to deliver money on such and such occasions? If a group of people creates an institution whose sole function is that I should give them money, I have, so far, no obligation whatever to give them money, because though they might have created what they think is a reason, it is not yet a reason for me. So, how can I use institutional reality to create desire-independent reasons for me?

It is at this point that we have to introduce the features of freedom and the first-person point of view. Our question now is, how can I create a reason for myself, a reason that will be binding on me in the future, even though I may not at that time have any desire to do the thing I created a reason for doing. I think the question becomes impossible to answer if you look at the phenomena from

the third-person point of view. From a third-person point of view, someone makes a bunch of noises through his mouth. He says, "I promise to wake you up at 6:00 A.M." How can his doing that ever create a reason that will bind his will? The only way to answer this question is to see, from the first-person point of view, what I think is going on, what I am trying to do, what my intention is when I make these sounds through my mouth. And once we see the matter from the first-person point of view, we can, I believe, see the solution to our puzzle. When I say "I promise to wake you at 6:00 A.M.," I see myself as freely creating a special type of desire-independent reason, an obligation, for me to wake you at 6:00 A.M. This is the whole point of promising. Indeed, that is what a promise is. It is the intentional creation of certain sort of obligation—and such obligations are by definition independent of the subsequent desires of the agent. But all I have said so far is that I made noises with certain intentions and that because I have those intentions, such and such seems to me to be the case. But how do we get from "it seems to be the case" to "it is the case," and to answer that question, we have to go back to what I just said about institutional structures. It is characteristic of these structures that *seems* is prior to *is*. If it seems to me that I am creating a promise, because that was my intention in doing what I did, and it seems to you that you have received a promise, and all of the other conditions (which I will not enumerate here but have enumerated in detail elsewhere),[9] if all the other conditions on the possibility of creating a promise are present, then I have created a

9. *Speech Acts: An Essay in the Philosophy of Language*, Cambridge: Cambridge University Press, 1969, chap. 3.

promise. I have intentionally created a new entity, which is binding on me in the future; it is a desire-independent reason for me, because I have freely and intentionally created it as such.

The ability to bind the will now creates a reason for the future act only because it is a manifestation of my freedom now. I said earlier that this shows why the slave doesn't have any reason to obey the slave owner, except desire-dependent reasons, even though both promissor and slave act within institutional structures. The only reasons the slave has are prudential reasons. The slave never exercised any freedom in creating a reason for himself to act. To see how within the institutional structure an agent can create external reasons for acting, it is essential to see that within the institutional structure, there is the possibility of the agent freely creating reasons for himself. There cannot be any question that it is a reason for him because he has freely and voluntarily created it as a reason for himself. Now, this is not to say, of course, that it is a reason that will override all other reasons. On the contrary, we know that in any real-life situation, there is likely to be a large number of competing reasons for any action, or against doing that action. When the time comes, the agent still may have to weigh his promise against all sorts of other competing reasons for doing or not doing something.

We have so far considered four features, time, institutional structures, the first-person point of view, and freedom. I now turn to the fifth: rationality. The ability to act rationally is a general set of capacities involving such things as the ability to recognize and operate with consistency, inference, recognition of evidence, and a large number of others. The features of rationality that are important for the present discussion involve the capacity

to operate in various ways with reasons for action. I want that to sound vague at this point because clarifying it is our next essential task.

Suppose I have freely acted with the intention of creating a desire-independent reason for me, suppose I have met all the conditions (on promising, or ordering a beer, or whatever), so that I really succeeded in creating that reason. Then, when the time comes, what do I need in order to recognize that there is such a reason? Assuming that I know all the facts, recognitional rationality is sufficient for acknowledging that the prior creation of a reason is now binding. The important thing is that you don't have to have some extra moral principle about promising or beer drinking in order to understand that the reason you created in the past as a binding reason for the present moment is precisely a binding reason in the present moment. It is sheer logical inconsistency to grant all the facts, about the creation and continuation of the obligation, and then to deny that you have a reason for acting.

VI Summary and Conclusion

In this chapter I have been concerned to show how human beings can create and be motivated to act on desire independent-reasons for action. What facts correspond to the claim that the agent has created such a reason, and what facts correspond to the claim that such a reason is a rational form of motivation to action? I have tried to discuss these questions at three levels. The first and most basic level is that of the atomic structure of the fundamental intentionality by which an agent can commit himself by imposing conditions of satisfaction on conditions of satisfaction. The second level is the level of

"phenomenology" where we discuss how it seems to the agent. The way it seems to the agent is that he is undertaking commitments through the free and intentional exercise of his will, in such a way as to bind his will in the future so that in the future he has a reason for an action that is independent of whether he desires to perform the action. And the third level is that of society in general—what are the social functions of having such systems of desire-independent reasons for action?

The basic facts that correspond to the claims that humans can create and be motivated to act on desire-independent reasons are these:

1. There must exist a structure sufficient for the creation of such institutional facts. These structures are invariably linguistic but they may involve other institutions as well. Such structures enable us to buy a house, order a beer, enroll in a university, etc.

2. Within these structures, if the agent acts with the appropriate intentions, that is sufficient for the creation of desire-independent reasons. Specifically, if the agent acts with the intention that his action should create such a reason, then if the circumstances are otherwise appropriate, he has created such a reason. The crucial intention is the intention that it be a reason. The reason does not derive from the institution; the institution provides only the vehicle for the creation of such reasons.

3. The logical form of the intentionality in the creation of such reasons is invariably the imposition of conditions of satisfaction on conditions of satisfaction. The purest case, so to speak, of the creation of a desire-independent reason for an action is the promise. Promising is, however, peculiar among speech acts in that it has the maker of the

promise as the subject of the propositional content, and has a self-referential component imposed on the conditions of satisfaction. The conditions of satisfaction of the promise are not only that the speaker do something, but that he do it because he made a promise to do it. There is, therefore, a self-referential component in promising, and this self-referential component does not exist in certain other sorts of speech acts. For example, it does not exist in assertions.

4. Once the obligation is created, it is a requirement of recognitional rationality that the agent should recognize it as binding on his subsequent behavior. The obligation has the structure of reasons for action. There is a factitive entity with the upward direction of fit, and the agent as subject.

5. Once a valid desire-independent reason for action has been created, that reason can motivate a desire to perform the action, just as the recognition of any other reason can motivate a desire to perform the action. To recognize a valid reason for doing something is already to recognize a valid reason for wanting to do it.

Appendix to Chapter 6: Internal and External Reasons

I have objected to Bernard Williams's claim that there are no such things as external reasons, that all reasons for an agent have to be internal to his motivational set. No doubt there are various objections one could make to this view, but the main thrust of my objection has been that there can be facts external to the agent's motivational set, such that rationality requires that the agent recognize these facts as reasons for action, even if there is nothing in his motivational set then and there that disposes him to recognize them as reasons. The two sorts of facts I have concentrated on are facts about long-term prudence and facts about the existence of desire-independent reasons such as obligations undertaken by the agent.

One last feature of the doctrine of internalism deserves special mention. There are interpretations of internalism on which the claim that there are no external reasons comes out as tautologically true, and I would not wish to be thought to be disagreeing with those. The problem is that the true tautological versions can easily be interpreted as substantive versions, which are false. (I am not suggesting that Williams himself makes this confusion.) And in this appendix I am going, all too briefly, to state the tautological versions and contrast them with the substantive versions.

The basic argument for internalism is that unless an agent has internal reasons, he would have nothing to reason *from*. An external reason, by definition, is one that is external to the agent, and consequently one he could not use to reason from. A corollary to this argument, and in a way the most powerful way of stating the argument, is that we could not explain an agent's actions in terms of his

reasons unless they were internal reasons, for only an internal reason can actually *motivate* the agent to act. So there are two closely related arguments for internalism, one about the process of reasoning, and one about motivation. Each of these admits of a tautological formulation, and of course I do not disagree with the tautological formulation.

Tautology A, reasoning: In order to reason in the mind on the basis of a reason, an agent has to have a reason in the mind to reason from.

The tautological version of a motivational thesis is as follows:

Tautology B, motivation: In order to be motivated by a reason in his mind, an agent has to have a reason in the mind that motivates him.

Both of these tautologies admit of a substantive reformulation that seems to me not tautological, but false. The substantive reformulation embodies the disagreement between the internalist and the externalist, where rationality is concerned.

Substantive thesis A: In order for any fact or factitive entity *R* to be a reason for agent *X*, *R* must already be a part of, or represented, in *X*'s motivational set *S*.

And the nontautological version of B is:

Substantive thesis B: All rational motivations are desires broadly construed, in the way that Williams describes *S*.

The substantive versions of internalism are immediately subject to counterexamples. Thesis A has the immediate consequence that facts about an agent's desire-independent

reasons for actions, such as facts about his long-term prudential interests and facts about his undertakings and obligations, cannot be reasons for action, even in cases where the agent is aware of these facts, unless the agent is disposed in his motivational set to act on these facts. Thesis B has the immediate consequence that at any point in an agent's life, and for any act type *T*, unless the agent right then and there has some desire, where desire is broadly construed, either to do an act of type *T*, or a desire for something such that there is a sound deliberative route from that desire to doing an act of type *T* as a means to satisfy the desire, then the agent has no reason to perform an act of type *T*. We have seen a number of cases where that is false, where the agent has a reason to perform an act even though these conditions are not satisfied.

So the dispute between the internalist and the externalist is about the existence of desire-independent reasons for action. The question is: are there reasons such that rationality alone requires the agent recognize them as motivations, whether or not they appeal to something in the agent's motivational set? According to the internalist, all reasons for action must be based on desires, broadly construed. According to the externalist, there are some reasons for actions that can themselves be the ground for desiring to do something, but are themselves neither desires nor based on desires. For example, I can have a desire to keep my promise because I recognize it as an obligation, without its being the case that the only reason I want to keep it is that I antecedently had a desire to keep all my promises.

Williams sometimes talks as if the recognition of an obligation already is an internal reason for action. But that claim is ambiguous. To say that *A* knows he has an obligation allows for at least two distinct possibilities.

1. *A* knows that he has an obligation, which he recognizes as a valid reason for acting and therefore as a reason for wanting to act.

2. *A* knows that he has an obligation, but he doesn't care a damn about it. Nothing in his motivational set inclines him to act on it.

Now the dispute between the internalist and the externalist comes out right here: for the externalist in both cases there are reasons for action. Indeed in both cases there are desire-independent reasons for action. For the internalist, only in case (1) is there a reason for action. Furthermore, according to the externalist, case (1) is misdescribed by internalism. The internalist thinks the recognition of a binding obligation as a valid reason is already a desire for action. The externalist thinks of it as the ground of a desire, which is itself a desire-independent reason for action.

In such cases it seems to me the defender of the internalist point of view might argue that the external reason can still function only if the agent has the capacity to recognize it as a binding obligation. And this leads to a third tautological version of internalism:

Tautology C: In the exercise of his internal dispositional capacities, in order for an agent to recognize an external reason as a reason, the agent has to have the internal capacity to recognize it as a reason.

But this is easily reinterpreted in a nontautological substantive version, which is false:

Substantive C: In order that any external fact can be a reason for an agent, the agent must have an internal disposition to recognize it as a reason.

It is easy to see how you can confuse the substantive with the tautological, but they are quite distinct. The tautological just says that in order to exercise a capacity the agent has to have the capacity. The substantive version says that nothing is a valid reason unless the agent is disposed to recognize it as such, and that, I have argued, is mistaken. It is part of the concept of rationality that there can be desire-independent reasons, reasons that are binding on a rational agent, regardless of desires and dispositions in his motivational set.