III | Convention Contrasted

1. Agreement

In Chapter I.3 we saw how agents involved in a single, unrepeated coordination problem might do well to achieve coordination by agreeing explicitly that each is to do his part of a certain coordination equilibrium. But I stated, too, that explicit agreement was not the only means of coordination.

Similarly, our explicit agreement to conform to some suitable regularity R is a means—a good means, but not the only one—of arranging once and for all to achieve coordination whenever a certain recurrent coordination problem S arises among us. In agreeing face to face to conform to R, each of us is manifesting his (pre-existing or newly formed) propensity to conform to R; and he is doing so under circumstances in which it is common knowledge among us that all of us are watching him. Consequently it becomes common knowledge that all of us are likely to conform to R. Since it is also common knowledge that each prefers to conform to R conditionally upon conformity by others involved with him in S,[1] each has all the more reason and propensity to conform—and this, too, is common knowledge. Common knowledge of a propensity to conform produces conforming action when S next arises; conforming action renews our common knowledge of a propensity to conform. A convention is under way.

Our convention is the product of our agreement and so—in a

[1] If it was not before, it becomes common knowledge during the discussion before we agree, as each of us manifests his conditional preference for conforming.

way—are all our conforming actions forevermore. But to say we act as we do because we once agreed to would be badly misleading. It suggests that our agreement continues to influence our actions directly, just as it did at first; actually its major effect is transmitted through a growing causal chain of expectations, actions, expectations, actions, and so on. The direct influence fades away in days, years, or lifetimes. We forget our agreement. We cease to feel bound by old promises (if our agreement *was* an exchange of promises; as we saw in the case of a single coordination problem, an exchange of manifestations of present intent is apt to be good enough). We leave the population, and are replaced by heirs who were not party to the agreement. But the indirect influence of the agreement is constantly renewed, and in time it comes to predominate. Then a convention created by agreement is no longer different from one created otherwise: it bears no trace of its origin.

In fact, a convention begun by agreement may not become a convention, on my definition, until the direct influence of the agreement has had time to fade. This depends on the nature of the agreement. Suppose we all swore a solemn and public oath to conform to R come what may. Then for a while we might all prefer *un*conditionally to conform to R, each determined that even were the others to break their oaths and conform to some alternative regularity R', still he would rather keep his oath. Even if our preferences for conforming to R were in fact conditional upon conformity by others, it might not be common knowledge that they were. For the onus of oath breaking might create uncertainty, be expected to create uncertainty, and so on. We have a convention only after the force of our promises has faded to the point where it is both true and common knowledge that each would conform to some alternative regularity R' instead of R if the others did.

If, on the other hand, we agreed by exchanging conditional promises binding us to conform to R only if others did, or by exchanging noncommittal declarations of intent, the resulting regularity would be a perfectly good convention at once.

To see how a convention might start by an agreement or otherwise,

let us take as our example a convention that does not yet exist: a convention among logicians establishing a standard notation. There is no such convention now. Since many notations are in use, everyone feels free to indulge his preference. But such a convention could exist. If any of today's notations were used as uniformly as standard arithmetical notation, everyone would use it. No one would insist on using his favorite notation if another notation were standard. (An eccentric notation would be more troublesome for readers used to a standard notation than it is for readers used to our present chaos.)

Imagine that the diversity of notations becomes more troublesome than it is now. Every author invents his own notation; some adopt it, some learn it but do not use it by choice, and others cannot even read it. Then the Association for Symbolic Logic might hold a meeting at which the problem is discussed; everyone present expresses the hope that a standard notation will be established; and *Principia* notation is elected by a vote of the meeting. If so, those who attended the meeting would consider themselves, and each other, to have expressed an intent to use *Principia* notation henceforth if most other logicians do too. Their exchange of declarations of intent in a face-to-face meeting would be an agreement sufficient to start a convention.

But the same convention might begin in other ways instead. Imagine some of the possibilities.

A few prominent logicians—say, the editors of the *Journal of Symbolic Logic*—might publish in the *JSL* a joint statement that the troublesome diversity of notation ought to be remedied, and that, in their opinion, *Principia* notation ought to be used exclusively henceforth. Since it is common knowledge among logicians that they all read the *JSL*, it would become common knowledge among logicians that most logicians had read the statement. And if it was common knowledge among logicians that most logicians would be inclined to follow such a suggestion if others did, then it would be common knowledge that most logicians would have a much increased propensity to use *Principia* notation. Given a sufficient interest in conforming to any standard, a convention to use *Principia* notation might result.

In this case only a few people play an active part in initiating the new convention; the rest are a responsive audience.

Or there might be no attempt to create a new convention. Several logicians, disturbed by the proliferation of notations and concerned for the intelligibility of their work, might decide independently to switch from their favorite notations to whichever notation seemed to be best known, namely, *Principia*. The resulting increase in use of *Principia* notation might seem like the beginning of a trend, so others would be inclined to switch and to expect each other to switch. Again the result would be a convention.

Or there might be a spate of works in *Principia* notation for no particular reason at all. Just by coincidence, many of the logicians who prefer *Principia* notation might happen to publish at once. An ostensible trend is created all the same, which others may follow.

In short, certain conditions—common knowledge of a general preference for using any sufficiently popular notation, plus common knowledge that logicians can tell how much the various notations are being used—tend somewhat to amplify any fluctuation in the logicians' expectations and propensities about their choice of notation. A convention is produced when a big enough fluctuation meets strong enough amplifying forces. The source of the fluctuation is unimportant, given its size. It does not matter whether it was created with the intention of starting a convention or whether it occurred in some or all of the population.

Granting that explicit agreement is only one of several possible origins for conventions, we may still wonder whether it enjoys some special status. Is it true, perhaps, that all conventions *could* originate by agreement? I offer three counterarguments.

First, recall the example of Hume's rowers, a conventional regularity we cannot describe. We cannot describe it in practice; if we could in principle, it would be by using more time and more measuring instruments than the rowers could have at their disposal. But this argument only goes to show that their convention could not originate by an agreement that is a purely verbal exchange. They can perfectly

well agree to row *thus*, specifying a rhythm of rowing by demonstrating it.

Next, suppose a convention produces coordination to serve some purpose that would be defeated somehow by the very act of agreeing. Then that convention could not originate by agreement, for an agreement would destroy the point of conforming to the convention. Take two people who find it expedient to keep up a facade of hostility (say, leaders of rival political parties). They could use a convention specifying what sort of opinions they are to profess on any topic, lest they find themselves in public agreement—to the embarrassment of both. But they cannot create a convention by agreeing. For that would destroy their facade by confessing their common interest in preserving it, leaving them nothing to coordinate for. (It is true that they might agree in secret. But secrecy would not help them if—to change the example—we suppose that their facade is one they present primarily *to themselves*. If so, however, the counterexample is suspect because the agents are deceiving themselves. They believe in their facade while taking precautions to avoid encountering the evidence that would disillusion them. So they cannot safely be treated as rational agents with coherent beliefs.)

Last, consider this argument, given by Quine and others.[2] The first convention of language to be established could not originate by an agreement conducted in a convention-governed language. So even if *any* convention of language could originate by such agreement, not *all* of them could. (Thus this argument differs from the two above, which purported to show that particular conventions could not originate by agreement.) I offer this rejoinder: an agreement sufficient to create a convention need not be a transaction involving language or any other conventional activity. All it takes is an exchange of

[2]By Quine in "Truth by Convention"; by Bertrand Russell in *The Analysis of Mind* (London: Allen and Unwin, 1921), p. 190; by William Alston in *Philosophy of Language* (Englewood Cliffs, New Jersey: Prentice-Hall, 1964), p. 57. All three take it to be an argument that there are no conventions of language, since they believe that conventions properly so-called must be created by agreement.

manifestations of a propensity to conform to a regularity. These manifestations might simply be displays of conforming action in various appropriate situations, done during a face-to-face meeting in order to create a convention. Such an exchange of displays might be called an "agreement" without stretching the term too far.

I take it that all three arguments are inconclusive. Construing "agreement" generously, maybe all conventions could, in principle, originate by agreements. What is clear is that they need not. And often they do not: Chapter I.5 should suggest familiar examples of conventions originating otherwise. Conventions are like fires: under favorable conditions, a sufficient concentration of heat spreads and perpetuates itself. The nature of the fire does not depend on the original source of heat. Matches may be our best fire starters, but that is no reason to think of fires started otherwise as any the less fires.

2. Social Contracts

It seems (subject to weak objections) that a convention is a regularity in behavior which holds *as if* in consequence of an agreement so to behave, by virtue of a general preference for general conformity to that regularity. Now this is just how one might describe a social contract, given the sophistication to treat the original making of the contract as a fictitious dramatization of our present reasons for conforming. Is my concept of convention nothing but our familiar concept of social contract, as inherited from Hobbes, Locke, and Rousseau, demythologized and applied to matters other than political allegiance and social solidarity?

It is not. The concept of social contract, as I understand it, is different in principle from that of convention (though there are descriptions, like the one above, crude enough to miss the difference). Nor do the extensions of these concepts coincide, though they overlap heavily.

I propose to define a *social contract* roughly as any regularity *R*

in the behavior of members of a population P when they are agents in a situation S, such that it is true, and common knowledge in P, that:

> Any member of P who is involved in S acts in conformity to R.
>
> Each member of P prefers the state of general conformity to R (by members of P in S) to a certain contextually definite state of general nonconformity to R, called the *state of nature* relative to social contract R.

The state of nature is not just any state of general nonconformity to R (by members of P in S); it is somehow distinguished.[3] The state of nature relative to Hobbes's social contract (whereby we constitute a leviathan by regular obedience to a sovereign) is understood to be anarchy and the war of all against all. It is not peaceful anarchy and not the existence of a leviathan under some other sovereign, although these would also be states of general nonconformity to our actual social contract. Peaceful anarchy does not qualify because Hobbes believes it to be unstable; the existence of a leviathan under some other sovereign does not qualify because it is too similar in kind to the social contract in question.

The state of nature relative to R is a state of general nonconformity to R; a state of general nonconformity to other regularities similar in kind to R; and a state in which no one is relying very heavily on any anticipated regularity in others' action. No one stands to lose too much if his expectations about his neighbors prove wrong. This is an especially stable state. And it is a state we might fall into if somehow we had to start from scratch with no strong mutual expectations (hence the *a priori* plausibility of the myth that primitive peoples live in a state of nature).

[3]See R. P. Wolff, "A Refutation of Rawls' Theorem on Justice," *Journal of Philosophy,* 63 (1966), pp. 179–190. Wolff, replying to the ideal contractualism proposed by John Rawls in "Justice as Fairness," *Philosophical Review,* 67 (1958), pp. 164–194, objects that Rawls has given us no way to identify the "baseline"—that is, the state of nature—with which a social contract should be compared.

Observe that because the state of nature relative to R is possible, it follows that members of P do not prefer unconditionally to conform to R—which, had it not been implied, should have been stated in the definition of social contract.

My definition of social contract paralleled that of convention as far as possible in order to show the location of difference: in the nature of the general preference for general conformity. Preferring something is preferring it *to* something else, and the second term of the preference is not the same. For convention, we require that each agent prefer general conformity to conformity by all but himself, ignoring his preferences regarding states of general nonconformity. For social contract, we require that each agent prefer general conformity to a certain state of general nonconformity, ignoring his preferences regarding conformity by all but himself.

Consider the regularity R of obeying a sovereign. Let us see how R might be a convention, a social contract, both, or neither. Suppose the status quo is this: we almost always obey the sovereign's commands:

(1) to refrain from taking one another's goods;
(2) To hand over all we can spare to support the sovereign in luxury;
(3) to help catch and punish anyone who breaks any of these three commands.

(If anyone is offended by this caricature of political society, let him turn it into an example about castaways, gangsters, or nations.) Each of us has some preference ranking of the following three states, each with its own advantages and disadvantages:

(SQ) The status quo. I am protected from my neighbors by the sovereign's power to enforce his first command. I am safe from the sovereign's police power because I do not try to disobey. But I am poor because I help to support the sovereign and I do not take others' goods.

(SN) The state of nature. Nobody commands general obedience, so I have no organized police power to fear. I do

not help to support a sovereign, for we have none. But I must work to protect my goods from my neighbors, and I live in fear that I will fail.

(*LD*) Lone disobedience. My goods are protected from my neighbors. I can get rich because sometimes I take my neighbors' goods and I do not contribute my full share toward the sovereign's support. But I live in fear that the sovereign, with my neighbors' help, will catch and punish me.

Ignoring indifference, each of us must have one of the six possible preference rankings, not necessarily the same for everyone:

SN	*SQ*	*SQ*	*LD*	*SN*	*LD*
SQ	*LD*	*SN*	*SQ*	*LD*	*SN*
LD	*SN*	*LD*	*SN*	*SQ*	*SQ*

(writing the most preferred state at the top, the least preferred one at the bottom). One's preferences will depend on his temperament; on his opinion of the character of his neighbors and of the sovereign; on his estimate of the likelihood that others would imitate his disobedience; on his estimate of his ability to make a living, to defend himself in the state of nature, or to escape punishment for disobedience; and on the values he assigns to security, wealth, relief from the task of self-defense, the welfare of his neighbors, the welfare of the sovereign, justice, peace, and trust.

If *R* is a convention, the only preference rankings that occur among us (except in a negligible minority) are the three in which *SQ* is ranked above *LD* so that each prefers to conform to *R* conditionally upon the others' conformity:

SN	*SQ*	*SQ*
SQ	*LD*	*SN*
LD	*SN*	*LD*

It makes no difference where *SN* occurs in anyone's ranking.

If *R* is a social contract, the only rankings that occur among us

(except in a negligible minority) are the three in which *SQ* is ranked above *SN* so that each is benefiting from the general conformity to *R*:

SQ	*SQ*	*LD*
LD	*SN*	*SQ*
SN	*LD*	*SN*

It makes no difference where *LD* occurs in anyone's ranking.

If *R* is both a social contract and a convention, the only rankings that occur among us (except in a negligible minority) are the two in which *SQ* is preferred to both alternatives:

SQ	*SQ*
LD	*SN*
SN	*LD*

If *R* is a convention but not a social contract, some or all of us have the preference ranking:

SN
SQ
LD

These people are trapped. They want the convention abandoned. But nobody dares to do his part of abandoning it unless he can count on many others to abandon it along with him. To take the extreme case: if *all* of us prefer the state of nature to the status quo, the convention owes its survival entirely to the difficulty of organizing *concerted* disobedience to the sovereign's commands. (If the sovereign values his position, he would do well to issue a fourth command prohibiting any efforts at such organization.) All conventions are metastable, but this one is apt to be less stable than most, since we have an incentive to get together to change it if we can. The convention not to admit that the emperor had no clothes was not a social contract. Everyone wanted to break it, but only the little boy dared to break it *alone*. When he did, the convention collapsed and the state of nature was restored.

If *R* is a social contract but not a convention, some or all of us have the preference ranking:

LD
SQ
SN

We may well ask why these people continue to conform. For in passing up opportunities to gain by disobeying the sovereign's commands, they must be acting against their own preferences.

If we think of someone's preferences as the resultant of *all* the more or less enduring forces that go into determining his choices, then action that regularly goes against preference is barely possible. It would have to be due to transient forces, and different ones on different occasions. And it is hard to see how it could become common knowledge that people would regularly act against preference, since action against preference is inherently exceptional. So on this, our most common, concept of preference, it is almost impossible for a social contract not to be a convention because some prefer not to conform although others do.

Sometimes, however, we think of preference more narrowly as the resultant of choice-determining forces *other than* a sense of duty. Our accepted moral obligations can and do regularly override our preferences in this narrow sense. So we could say of a man that his preferences are

LD
SQ
SN

but that he obeys the sovereign's commands against his own preferences because he considers himself to be under a moral obligation to do so. He might believe with Locke that he gave his fellowmen a tacit promise to obey if they obeyed, when he first passed up an opportunity to leave the country; or he might believe with John Rawls that he is under an obligation of fair play to reciprocate the

benefits he has willingly received through the others' obedience.[4] Either way, his obligation arises because he prefers the status quo to the state of nature. Hence, for any social contract, the conditions of such obligations are present for everyone. If everyone will recognize such obligations, everyone will honor the social contract whether or not he prefers to. The social contract will persist, and it may be common knowledge that it will. But it is not a convention, not if the preferences mentioned in the definition of convention are taken as preferences in the narrow sense.

If we return to our ordinary, wider concept of preference, it remains true that many social contracts will be sustained by the moral obligations of tacit consent or fair play, as recognized by the agents involved. But these accepted obligations will be counted as a component of preferences, not as an independent choice-determining force. Since we accept these obligations, our preference rankings will be

$$\begin{matrix} SQ \\ LD \\ SN \end{matrix} \quad \text{or} \quad \begin{matrix} SQ \\ SN \\ LD \end{matrix}$$

instead of

$$\begin{matrix} LD \\ SQ \\ SN \end{matrix}$$

as they might be (for some of us) otherwise. So our social contract is a convention after all. But it is a convention because of the modification of our preferences by obligations, and these obligations exist because it is a social contract.

Thus the possibility of a social contract that is not a convention

[4] But the content of our illustrative social contract is not the content either Locke or Rawls had in mind; rather it is the simple and desperate contract of Hobbes's *Leviathan*. I see no reason why men should not adhere to Hobbes's contract for Locke's or Rawls's reasons.

(in the way so far considered) is problematic; it depends on adopting the less common of our two concepts of preference, that in which preference is opposed to obligation.

There is a quite different way, less problematic, for a social contract not to be a convention. Recall that we required a convention to be one of several alternative conventions, whereas a social contract need have no other alternative than the state of nature. The state of nature need not be a state in which we achieve coordination equilibria conforming to a regularity. Certainly Hobbes's state of nature is not: battles in the war of all against all do not result in equilibria (since the loser will wish he had adopted a different strategy), let alone coordination equilibria.[5] So a social contract may fail to be a convention for lack of an alternative, even though it is a regularity whereby we reach coordination equilibria and everyone therefore prefers to conform to it if the others do.

If we recurrently find ourselves involved in a certain situation (not a coordination problem) represented by the payoff matrix in Figure 30, and all of us regularly act so as to achieve the coordination

	C1	C2	C3
R1	1 1	.5 −.5	.5 −.5
R2	−.5 .5	1 −1	−1 1
R3	−.5 .5	−1 1	1 −1

Figure 30

[5] In Rousseau's example of the stag hunt, on the other hand, the state of nature relative to the social contract of all helping to hunt the stag together is the state in which we all catch rabbits by ourselves; and this state of nature *is* a state in which we reach coordination equilibria.

equilibrium ⟨*R1*, *C1*⟩, our regularity is a social contract, and it is a regularity whereby we achieve coordination equilibria, but it is not a convention. For there is only one other stable state, the state of nature: that in which we choose at random every time between *R2* and *R3* (or *C2* and *C3*). And the outcomes we reach in the state of nature—⟨*R2*, *C2*⟩, ⟨*R2*, *C3*⟩, ⟨*R3*, *C2*⟩, and ⟨*R3*, *C3*⟩—are not coordination equilibria, indeed not equilibria at all. (We do achieve an equilibrium combination of *mixed* strategies, but that is an equilibrium only in an extended sense and still is not a coordination equilibrium.) So the state of nature is not a convention, and our regularity of choosing *R1* and *C1* cannot be a convention either.

As an example, suppose that every Friday the same ten of us go to a Chinese restaurant where we are served, among other things, a plate of twenty fried shrimp. We would like three or four each, but for the sake of good relations each is willing to limit himself to two if and only if everyone else limits himself too. No one cares to restrain himself unless he can be sure that thereby he allows everyone to get his share. So there are two stable states: a social contract, whereby everyone takes two, or the state of nature, a scramble in which the first comers take all they want until none are left. The state of nature is not a state in which we achieve coordination equilibria, so the social contract is not a convention, although each prefers to conform to it if the others do. And that is as it should be: this social contract does seem to lack the characteristic arbitrariness of a convention.

Finally, a social contract might fail to be a convention, or vice versa, in still another way: the items of common knowledge required in the definition of convention are not the same as those required in the definition of social contract. So a suitable lack of common knowledge might disqualify a regularity as a convention but not as a social contract, or vice versa. It is hard to see how such a discriminating lack of common knowledge could occur; generally, the conditions that make for common knowledge of any of the facts of a case make for common knowledge of all of them.

3. Norms

The definition I gave of convention did not contain normative terms: "ought," "should," "good," and others. Nor have we reason to expect normative terms to occur essentially in any equivalent definition. So "convention" itself, on my analysis, is not a normative term.

Nevertheless, conventions may *be* a species of norms: regularities to which we believe one ought to conform. I shall argue that they are. There are certain probable consequences implied by the fact that an action would conform to a convention (whatever the action and whatever the convention) which are presumptive reasons, according to our common opinions, why that action ought to be done.

Suppose R is a convention in population P regarding behavior in situation S. And suppose I am a member of population P in situation S. Then by the definition of convention, and without regard to what R, P, and S may be, this supposition makes it probable that:

(1) Most other members of P involved with me in situation S will conform to R.
(2) I prefer that, if most other members of P involved with me in S will conform, then I conform also.
(3) Most other members of P involved with me in S expect, with reason, that I will conform.
(4) Most other members of P involved with me in S prefer that, if most of them conform, I conform also.
(5) I have reason to believe that (1)–(4) hold.

Had we supposed R to be a convention to degrees 1, 1, 1, 1, 1, 1, then that and the supposition that I am a member of P in S would have implied (1)–(5), even strengthened by putting "all" for "most" throughout. But, since we chose to allow for conventions that are less than perfectly conventional, we must hedge by stating (1)–(5) to allow exceptions and by allowing exceptions to (1)–(5) themselves. But (1)–(5) must hold in *most* cases in which people decide whether

to conform to any convention. Thus in any given case—barring evidence that the case is exceptional—each of (1)–(5) probably holds: so probably (although less probably) all of them hold.

And if they all do hold in any case, then so do these:

(6) I have reason to believe that my conforming would answer to my own preferences.

(7) I have reason to believe that my conforming would answer to the preferences of most other members of *P* involved with me in *S*; and that they have reason to expect me to conform.

And (6) and (7), when true, are presumptive reasons why I ought to conform. For we do presume, other things being equal, that one ought to do what answers to his own preferences. And we presume, other things being equal, that one ought to do what answers to others' preferences, especially when they may reasonably expect one to do so. For any action conforming to any convention, then, we would recognize these two (probable and presumptive) reasons why it ought to be done. We would not, so far as I can tell, recognize any similarly general reasons why it ought not to be done. This is what I mean by calling conventions a species of norms.

Of course, for any particular action conforming to a convention, there may be all sorts of other reasons why it ought or ought not to be done. (There might also be reasons to believe that the case is an exceptional one in which [6] or [7] does not hold.) The convention in question might be one that was adopted by an exchange of promises, so that a conforming action ought to be done to keep one's promise. Or it might be a convention that is also a social contract, so that a conforming action ought to be done to reciprocate the benefit one derives from conformity by others. On the other hand, it might be an understanding between oligopolists to fix prices or between pickpockets to work in a team, so that a conforming action ought not to be done since it is not in the public interest. As always, the various presumptive reasons why an action ought or ought not to be done are balanced off against each other; (6) and (7) might

easily be outweighed. Their importance is not that they are especially weighty considerations. Rather, it is that they are especially general: for any convention whatever, they must enter into deliberations about whether to conform to it.

Any convention is, by definition, a norm which there is some presumption that one ought to conform to. I shall now argue that it is also, by definition, a socially enforced norm: one is expected to conform, and failure to conform tends to evoke unfavorable responses from others. For if R is a convention in population P regarding behavior in situation S, and I am a member of P in S, then (by the definition of convention, and no matter what R, P, and S may be) probably:

(8) Most other members of P involved with me in S expect me to conform.

(9) Most other members of P involved with me in S have reason to believe that conditions (1)–(5) hold.

And whenever (9) holds, so do these:

(10) Most members of P involved with me in S have reason to believe that my conforming would answer to my own preferences.

(11) Most members of P involved with me in S have reason to believe that I have reason to believe both that my conforming would answer to their preferences and that they have reason to expect me to conform.

So if they see me fail to conform, not only have I gone against their expectations; they will probably be in a position to infer that I have knowingly acted contrary to my own preferences, and contrary to their preferences and their reasonable expectations. They will be surprised, and they will tend to explain my conduct discreditably. The poor opinions they form of me, and their reproaches, punishment, and distrust are the unfavorable responses I have evoked by my failure to conform to the convention.

Consider our example of a convention that the original caller calls back and the called party waits when a telephone call is cut off. Suppose I fail to conform—I wait when I am the original caller or I call back when I am the called party. My partner knows what I did. He knows that I should have known that by acting as I did I would fail to restore our connection. He may guess that I acted without thinking; or that I was too dull ever to learn the convention; or that I was bored with talking to him anyway and did not want the call restored; or that I expected *him* to be at fault in one of these ways and violated the convention to counteract the violation I expected from him. Whatever he thinks, his opinion of me suffers. So does the way he is likely to treat me in the future. These are bad consequences, and my interest in avoiding them strengthens my conditional preference for conforming.

4. Rules

We would certainly call many conventions rules. For instance, those presented in Chapter I.5 as established solutions to the eleven sample coordination problems might all naturally be called rules—though probably with a qualification: "tacit" rules, "informal" rules, "unwritten" rules, or the like. But certainly not all so-called rules are conventions. Let us consider several kinds of counterexamples.

Sometimes mere generalizations, laws of nature, or even mathematical truths are called rules. These rules may have nothing to do with the conduct of human agents, except that human agents might benefit by taking account of them. For instance, we have this passage in a cookbook:

> Here is a cardinal rule that has very few exceptions: *All* meat is more tender and juicy if cooked at *low* instead of high temperature.[6]

[6]B. B. McLean and T. W. Campbell, *The Meat and Poultry Cookbook* (New York: Pocket Books, 1960), p. 19.

And this theorem that an algebra text calls "Descartes' Rule of Signs":

> An equation $f(x) = 0$ cannot have more positive roots than there are changes of sign in $f(x)$, and cannot have more negative roots than there are changes of sign in $f(-x)$.[7]

Other so-called rules are strategic maxims, hypothetical imperatives stating what a human agent might do to gain some end. These rules state generalizations regarding the tendency of certain actions to accomplish certain ends. We are given these "general rules for using any insecticide":

> Treat any household insecticide, no matter how labeled, as a poison. Never use insecticides on nursery walls, playpens, cribs, toys, or places where infants creep. Buy insecticides only as needed. Store them in a locked cabinet. Once insects are conquered, bury all leftover insecticides deep in the trash container.[8]

And this "rule for reducing a recurring decimal to a vulgar fraction" (a hypothetical imperative based on a theorem):

> For the numerator subtract the integral number consisting of the non-recurring figures from the integral number consisting of the non-recurring and recurring figures; for the denominator take a number consisting of as many nines as there are recurring figures followed by as many ciphers as there are non-recurring figures.[9]

Here are rules for cultivating taste:

> the first rule for the student of wine is to trust his own palate—to believe in its physical ability to record as many sensations as anybody else's . . . The second important rule for the student

[7] H. S. Hall and S. R. Knight, *Higher Algebra* (London: Macmillan, 1960), p. 459.
[8] *Consumer Reports,* 30 (1965), no. 12, p. 156.
[9] Hall and Knight, *Higher Algebra,* p. 43.

> is that he must have the courage to change his mind. As experience grows and perception becomes keener, his taste is certain to change and wines which at first pleased may now bore or actively displease.[10]

Other rules are hypothetical imperatives reinforced by authoritative codification and enforced by sanctions. This rule might appear on a poster in a chemical plant:

> Employees are not to smoke within 100 yards of any acetone vat; violation will be considered ground for immediate dismissal.

Not to smoke near an acetone vat already answers to any worker's interest in avoiding fires, regardless of whether or not his fellow workers smoke near acetone vats and regardless of whether management has laid down a rule against it. The official rule reminds employees of this hypothetical imperative. It also notifies them of the fact that it is management's policy to fire anyone caught smoking near an acetone vat; it is a threat or warning[11] designed to deter them further from doing so.

Other rules are threats or warnings issued by some authority or power to control the behavior of a class of people against their own preferences. A POW camp might have a rule that prisoners are not to gather in groups of more than six, violation to be punished by ten days on bread and water. A local protection mob might make a rule that lunch counters are to rent one pinball machine for every ten seats, violation to be punished by arson. Since one's incentive to obey is the same whether or not the rest obey (unless mass disobedience would destroy the enforcer's will or power to punish), a rule of this kind is not a convention.

[10] Allan Sichel, *The Penguin Book of Wines* (Baltimore: Penguin Books, 1965), p. 22.

[11] Most likely a warning. The distinction is Schelling's; see *Strategy of Conflict,* pp. 123–124. A *warning* is a statement that if you do *A* which I don't like, you probably will thereby give me a good reason to do *B* which you don't like. A *threat* is a declaration of my present intent to do *B* if you do *A*, whether or not I would then have any good reason to do *B*.

It might, however, be conventional in the sense that it forms the *content* of a convention among the rule makers. Suppose the commandant of a POW camp who is exceptionally strict or lenient will get into trouble. Then all the commandants have an interest in adopting the same schedule of penalties. If so, it might be a convention among the commandants that the penalty for gathering in groups of more than six is to be ten days on bread and water. The rules assigning that penalty in the several camps are, in a sense, conventional; but they are not conventions.

Other rules codify regularities of the kinds discussed in section 2 as social contracts that are not conventions. If we adopt the narrow concept of preference in which one's preferences do not include his acceptance of moral obligations—for example, obligations of tacit consent or fair play—this class of rules is a large one. These rules prescribe behavior for each agent which may go against his own preferences (in the narrow sense) but which answers to the preferences of everyone else concerned. They are hypothetical imperatives stating what one should do to keep tacit promises, what one should do to reciprocate benefits, and so on. They may also carry threats or warnings that violation will incur specified (institutionalized or informal) sanctions. Criminal law may consist largely of rules of this sort, at least in a traditional society without legislation. So may part of our social morality, for instance the rule that one should keep his promises. So may a library's regulations regarding the return of borrowed books.

Finally, there can be rules that are not conventions only because they are enforced with sanctions so strong that one would have a decisive reason to obey even if others did not. Take the convention whereby logicians might establish a standard notation; and suppose that after the convention had existed for twenty years, any editor would reject out of hand a manuscript using nonstandard notation. The editors would then have made a rule requiring standard notation, a rule enforced by the sanction of nonpublication. But it is no longer a convention, since each logician has a decisive reason to use standard notation whether his colleagues do or not. He still

wants to use whatever notation his colleagues use, and he would like to follow them if they all switched; but he would not be likely to care enough to forgo publication. (I assume some inertia on the part of the editors, so that they might try to insist on the old standard notation for a time, even after most logicians had switched to a new notation. If the editors too would switch immediately, their enforcement of standard notation does not detract from its conventionality.) Under these conditions, it is not true of any nonstandard notation that everyone would prefer to use it if the others did; therefore the use of standard notation is no longer a convention.

Perhaps there are other kinds of so-called rules that are not conventions, but this completes my list.

It is harder to argue that some conventions are not naturally called rules. (Indeed, it is hard to show that there is *any* regularity that could not be called a rule in *some* context.) But consider conventions that coexist and contrast with rules that are rules par excellence, say in a game. The game of Jotto is definable as activity conforming to the following rules.

> There are two players.
>
> Each chooses a five-letter word for the other to guess.
>
> Each in turn proposes a five-letter test word of his choice, and the other tells him how many letters his word has in common with the word he is guessing.
>
> When a player proposes the word the other chose for him to guess, he wins.

These are perfect specimens of rules. They define the game of Jotto. They are easily codified, as above, and their codifications are used in teaching people to play Jotto. Their violation would be taken as decisive evidence of inability or unwillingness to play Jotto. They are conventional; but they are not the only conventions in the game. Any group of players will develop understandings—tacit, local, temporary, informal conventions—to settle questions left open by the listed rules. What foreign words, slang, proper names, acronyms, or coinages are admissible words? May a player have an earlier

answer repeated (without wasting his turn) if he thinks a mistake was made? And so on. We might call these understandings rules—unwritten rules, informal rules—if we like; but we would also be inclined to emphasize their differences from the listed rules by saying that they are not rules, but only conventions.

I hope my examples have left an impression that the class of so-called rules is a miscellany, with many debatable members. We might be tempted to try distinguishing several senses of the word "rule," hoping that one of them would agree with my definition of convention. I doubt that the project would succeed. Many senses could be proposed, but probably they would turn out not to be distinct enough to merit the name of different senses. We seem to be dealing with an especially messy cluster concept, and one in which the relative importance of different conditions varies with the subject matter, with the contrasts one wants to make, and with one's philosophical preconceptions. (It should be clear now why I have been contrasting conventions not with rules but with "so-called rules." I wanted to recognize the variation in what we would call a rule without saying whether this variation is ambiguity, boundary vagueness, or what.)

William Alston—speaking, I think, for many philosophers—has made the following proposal.

> Like the social contract theory in political science, the idea that words get their meaning by convention is a myth if taken literally. But like the social contract theory, it may be an embodiment, in mythical form, of important truths that could be stated in more sober terms. It is our position that this truth is best stated in terms of the notions of rules. That is, what really demarcates symbols is the fact that they have what meaning they have by virtue of the fact that for each there are rules in force, in some community, that govern their use . . . Henceforth, we shall feel free to use the term "conventional" purged of misleading associations, as shorthand for "on the basis of rules."[12]

[12] *Philosophy of Language,* pp. 57–58.

But if my analysis of convention is sound, and if the class of so-called rules is as miscellaneous as my examples seem to show, then it would be better to do exactly the opposite: to understand the "rules of language" we encounter in the works of philosophers of language as tacit conventions. We have no excuse for being misled by the misleading association of convention with explicit agreement.

Nor shall we be misled by misleading associations of the word "rule." When someone says that language is governed by rules, we are likely to think of rules that have been codified by some authority, or easily could be; of rules that are enforced by sanctions, formal or informal; of rules that are mentioned in teaching or criticizing the use of language. Paul Ziff, for one, has been misled into skepticism by just these misleading associations, as we see from his denunciation of rules of language:

> I am concerned with regularities: I am not concerned with rules. Rules have virtually nothing to do with speaking or understanding a natural language.
>
> Philosophers are apt to have the following picture of language. Speaking a language is a matter of engaging in a certain activity, an activity in accordance with certain rules. If the rules of the language are violated (or infringed, or broken, etc.) the aim of language, viz. communication, cannot save *per accidens* be achieved. Rules are laid down in the teaching of language and they are appealed to in the course of criticizing a person's linguistic performance.
>
> The picture admits of variation, of elaboration, but I shall not probe deeper into these mysteries. Such a picture of language can produce, can be the product of, nothing but confusion. An appeal to rules in the course of discussing the regularities to be found in a natural language is as irrelevant as an appeal to the laws of Massachusetts while discussing the laws of motion.[13]

[13] *Semantic Analysis* (Ithaca, New York: Cornell University Press, 1960), pp. 34–35.

But if we take the philosophers' rules of language to be tacit conventions of language, we escape Ziff's attack. For we are not supposing that these so-called rules are laid down in the teaching of language—we need not even suppose that language is taught—and we are not supposing that these so-called rules are appealed to in the course of criticizing linguistic performance. Yet our rules will not just be regularities in verbal behavior; they will be regularities in verbal behavior, and in expectations and preferences regarding verbal behavior, and in expectations regarding these expectations and preferences, and so on.

5. Conformative Behavior

David Shwayder, in *The Stratification of Behaviour,*[14] undertakes to give an analysis of a concept of rule. He introduces a correlative concept for the purpose: that of *conformative behavior.* The term is defined explicitly without mentioning rules. Then Shwayder's concept of rule—our ordinary one, he hopes—is defined explicitly in terms of conformative behavior. The two concepts are to be related thus: rules are certain facts that can be reasons for an agent's behavior; "conformative behavior is of a kind which requires that the agent have a certain kind of reason or mistaken reason, which we can schematize as 'That's the rule'" (p. 238). This includes deliberately *not* conforming to a supposed rule; conform*ative* behavior is not limited to conform*ing* behavior.

Shwayder's doctrine of rule and conformative behavior closely resembles my analysis of convention. (Shwayder happens also to use the term "convention," but for something quite different. Since his "convention" is much like H. P. Grice's "meaning$_{nn}$" which I shall discuss in Chapter IV.5, I shall not consider it further.) Behavior that qualifies under my definition as conformity to a convention qualifies under Shwayder's as conformative behavior, and for much the same reasons. But some differences emerge. Conformative behavior in-

[14] New York: Humanities Press, 1965.

cludes conformity to some regularities that are not conventions and that can scarcely even be called rules.

Shwayder leads up to his definition of rules and conformative behavior by stating several theses about rules which ought to follow from any satisfactory theory, and which will follow from his.

(1) Conformative behavior is not merely behavior that happens to conform to rules. "The agent must himself either conform to or act in violation of the rule. A condition for that is his believing that a rule exists" (p. 241).

(2) Some rules must be formulated in advance; others need not be. But in either case the rule itself is distinct from a formulation or statement of that rule.

(3) Rules are not, or not merely, regularities in behavior. Still less are they the generalizations we may frame about regularities in behavior. But regularities in behavior may be due to the existence of rules to which agents regularly conform.

(4) Rules are primarily rules of a community of agents. Conformative behavior presumes community rules. Private rules are possible, but they are somehow secondary to community rules and it is essential that they are capable of becoming community rules.

(5) Rules are certain reasons for acting: they are facts of a certain sort whose supposed existence is a reason for certain behavior, namely, conformative behavior. (This is so even in nonstandard cases in which the agent is mistaken about the existence of a rule or is acting to violate a rule.) A rule is a reason for acting by virtue of some such principle as "one ought to act to conform to rule," "one ought to act to avoid the penalties or discomforts which may ensue upon an infraction or breach of the rule," and so on (p. 251).

Now we reach Shwayder's central thesis about the nature of rules: "Community rules are systems of expectation. An agent conforms to such a rule if he acts for the reason that members of the community are entitled to expect him so to act" (p. 252). He elaborates:

> Confining ourselves to community rule, the idea is this: One follows a rule if he conforms to what he sees are the legitimate

> expectations of others; and the existence of a rule is, moreover, what entitles the others to their expectations, thus rendering them "legitimate." A community rule exists if the members of a community regulate their affairs according to what other members of the community would legitimately expect them to do. The rule is at once the expectations one conforms to and what legitimizes or warrants those expectations. The rule is, as it were, a system of community, mutual expectation. When one conforms to a rule he acts in the knowledge or belief that others would expect him so to behave. That the others are entitled to those expectations is his reason. Of course one may act in violation of such rules; but even there too one must believe that others have legitimate expectations. If one has no thoughts about what is expected of him, then he can neither conform to nor act in violation of the rule. (p. 253)

But Shwayder has not yet defined conformative behavior to his satisfaction, for he is still mentioning the rule itself. He must find some way to restate his offending condition that "the existence of a rule is . . . what entitles the others to their expectations, thus rendering them 'legitimate.'" It should be the existence of some *feature* of the rule—one that can be otherwise described.

Before Shwayder goes on, though, he temporarily restricts himself to "what is surely the most fundamental and persistently the most common case, that of rules of community, with the additional conditions (1) that some of the members of that community are *present*, where (2) the agents act to *conform*, and (3) the agent and his observers may be taken to *know* all that is relevant to be known about the situation" (p. 254). (The "observers" are extraneous describers of behavior whose concepts Shwayder is examining. So condition [3] is not a common-knowledge condition, and Shwayder is vulnerable to the arguments that led us to establish one.)

Conformative behavior in the fundamental case has so far been described, in effect, as behavior justifiable by the reasoning represented in Figure 31. (Where Shwayder speaks of knowledge, I con-

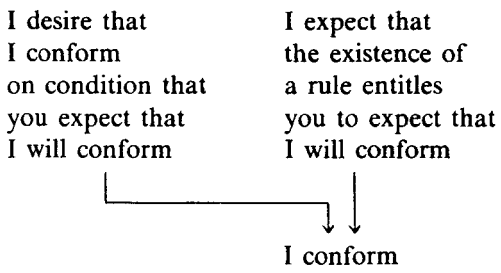

Figure 31

tinue to speak of expectations. But the difference does not matter, since in any normal case the expectations will also be knowledge.) The problem is to redescribe this piece of reasoning in a way that does not mention the rule itself.

Shwayder's solution is given in his formula: "I act from the knowledge that others know that I will act from the knowledge that they expect me so to behave" (p. 256). I take this to mean that there must be part of the agent's justification which is represented by Figure 32. Combining these two fragments of the agent's justification,

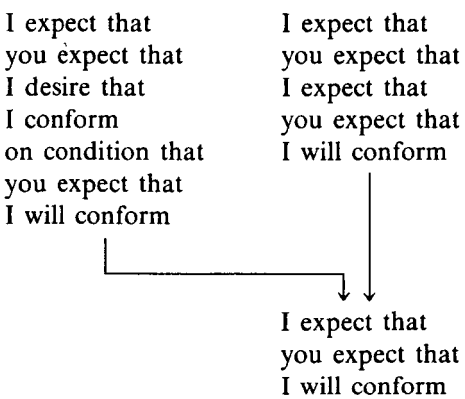

Figure 32

we find that his justification must be represented in part by Figure 33. I take it that an action meets Shwayder's definition of conformative behavior if and only if it can be justified by reasoning

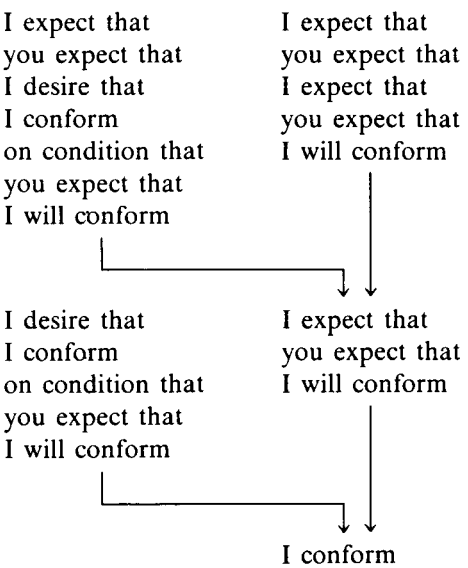

Figure 33

that fits this schema, which we may call the *schema for conformative behavior.*

Shwayder has made the following changes in his condition for conformative behavior. On his first version, I know that the others are entitled to their expectation about my behavior by "the existence of a rule." On his second version, I know that the others are entitled to their expectation about my behavior because they could acquire that expectation by deriving it from their knowledge that I know they expect me so to behave, together with their knowledge that I will try to do what is expected of me. Moreover, it is by replicating just this derivation that *I* could obtain my own knowledge of their expectation about my behavior.

By this change, Shwayder has succeeded in analyzing out his residual mention of the rule. Now that his definition of conformative behavior in the fundamental case is satisfactory, he is free to go on, without circularity, to define *rules* as those mutual expectations about behavior which figure in conformative behavior as reasons for acting.

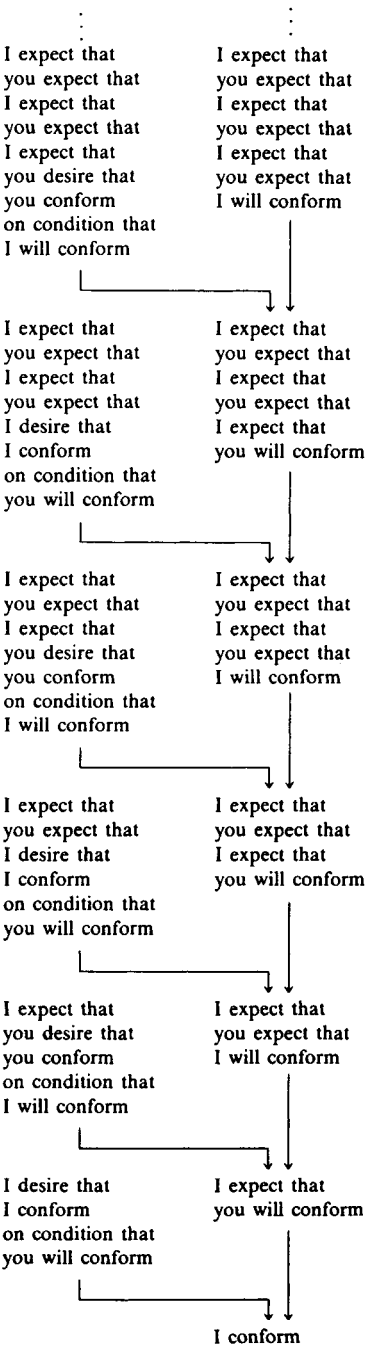

Figure 34

They are just those expectations about behavior that the other members of an agent's community have derived in the specified way.

Next Shwayder extends his definitions by relaxing his restrictions on the fundamental case. He defines conformative behavior in which no other members of the agent's community are present; in which the agent is acting to violate a rule; in which the agent is mistaken about the existence of a rule; or in which the rule involved is a private, potentially public rule. But we will not pursue these extensions. For it is Shwayder's doctrine of conformative behavior in the fundamental case that comes closest to my analysis of conformity to convention.

When we take any example of an action conforming to a convention, on my analysis, it will also be found to satisfy Shwayder's definition of conformative behavior in the fundamental case. That is no accident. Whenever I conform to a convention, my action is justifiable by replications; the depth of nesting is limited only by the availability of ancillary premises regarding rationality. Taking the two-person case, for simplicity, and ignoring the use of rationality premises, my justification may be represented by our usual replication schema, as shown in Figure 34. But there is a different way to represent essentially the same justification of my action. Consider any two consecutive stages of the above schema. Together they take me from an $(n + 2)$th-order expectation about action to an nth-order expectation about action (or, if $n = 0$, a decision to act) via an intermediate $(n + 1)$th-order expectation about action. But the same reasoning—same premises, same conclusion—could be carried out in a different order, with a different intermediate step. The last two stages, for instance, could be rearranged as shown in Figure 35. The same premises lead me to the same action, but the premises are used in a different order. There is a new intermediate step—I derive the desire to conform on condition that you expect me to. Given that, my expectation that you will conform becomes superflous. Rearranging each pair of stages from the bottom up, we obtain the *rearranged replication schema* shown in Figure 36. And now we can

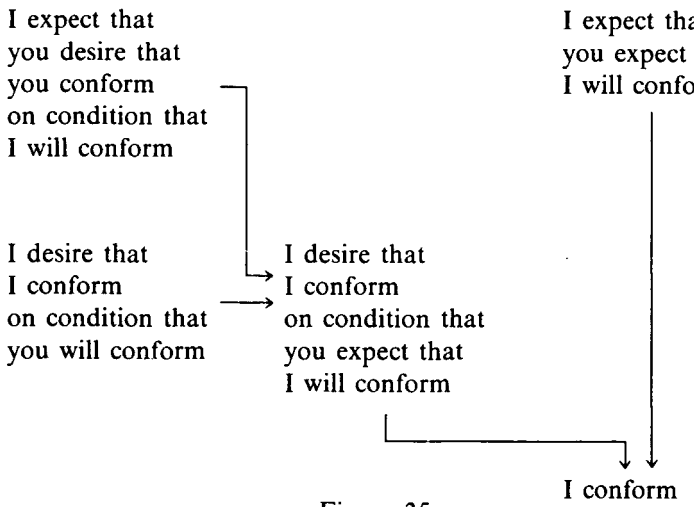

Figure 35

see that the schema for conformative behavior is just the central part (within the boundary) of the rearranged replication schema. It follows that any action in conformity to a convention, being justifiable by replications and hence by rearranged replications, meets Shwayder's definition of conformative behavior.

The converse does not hold, however. Suppose I want to conform if others expect me to, but without regard to anything *they* will do *because* they expect me to conform. Then the justification of my action cannot fit the rearranged replication schema. Since the left-hand column will be missing, my action cannot be conformity to a convention. But my justification might still fit the schema for conformative behavior. If so, my action would be conformative behavior. I would be the only *agent* in the situation; the others would be involved merely as supposed holders of expectations about me. I might just be trying to prove to myself that I can live up to others' expectations if I want to. I might want to get them to think me a dull, predictable fellow. I might want to avoid disappointing or surprising them. In any of these cases, I may behave as I do "from the knowledge that others know that I will act from the knowledge

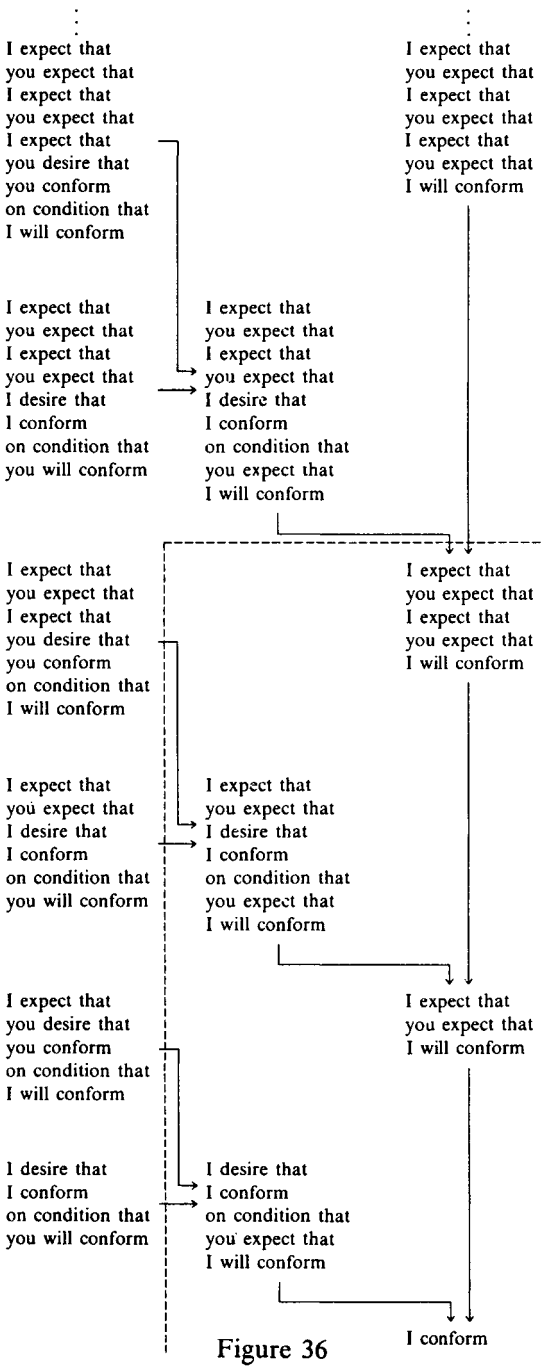

Figure 36

that they expect me so to behave." But these cases of conformative behavior seem to be cases Shwayder did not intend. They do not fit his formula, "I act for the reason 'That's the rule.'" I *do* act for the reason "That's their legitimate exception"—and the legitimacy is of the right sort—but their legitimate expectation in such a case cannot possibly be called a rule.

Shwayder has gained some generality by requiring only a central fragment, not the whole, of the rearranged replication schema. But it is not clear why this generality is desirable. And to buy it, he has left out an important fact about the intended sort of conformative behavior: namely, that I want to conform to your expectations *because of* the way I expect you to act on your expectations. If I thought you would not act on your expectations, I would concern myself with how you would act, not with what you expect. When this fact is left out of the story, our understanding of the phenomenon is badly distorted.

There is a second way for an action to be conformative behavior without being conformity to convention, even if the justification of the action does fit the whole rearranged replication schema. Justification by replications, and hence justification by rearranged replications, applies to *any* action that is part of a proper equilibrium. But the equilibrium does not have to be in a coordination problem, and it does not have to be a coordination equilibrium. It might even be a unique equilibrium in a game of pure conflict.

Suppose that for some reason pairs of us must often play the game of penny matching with an option of "calling off," a game represented by the payoff matrix shown in Figure 37. This is a game of pure conflict: Row-chooser and Column-chooser must play simultaneously by calling off the game (*R1* or *C1*), by putting down a penny head up (*R2* or *C2*), or by putting down a penny tail up (*R3* or *C3*). If the pennies match, Column-chooser takes them both. If they fail to match, Row-chooser takes them both. If both players call off, each keeps his penny. If only one calls off, the other pays him a halfpenny. There is a unique equilibrium ⟨*R1*, *C1*⟩; it is not a coordination equilibrium.

	C1	C2	C3
R1	0 0	−.5 .5	−.5 .5
R2	.5 −.5	1 −1	−1 1
R3	.5 −.5	−1 1	1 −1

Figure 37

I should call off unless I am quite confident that you will play heads (or tails), in which case I should play heads or tails (as appropriate) to beat you. In particular, I prefer to call off if you will. If we play often, then once the nature of the game has become common knowledge between us, we call off every time. It is common knowledge between us that we call off, and that we do so because we expect each other to.

When I call off, my action is conformative behavior. Its justification fits the rearranged replication schema and *a fortiori* the schema for conformative behavior. I am acting from the knowledge that you are entitled to expect me to call off, or—more precisely—that you know I will act from the knowledge that you expect me to call off.

But our regularity of calling off does not meet the definition of convention. Its exclusion is justified by two important differences between it and clear cases of convention: (1) The players' equilibria are not coordination equilibria. They are not cooperation in the common interest, but deadlocked compromises between completely opposed interests. Neither is satisfied with $\langle R1, C1 \rangle$. Neither could have achieved a better outcome by acting otherwise himself, but each wishes the other had acted otherwise. (2) The players' equilibria are the only possible ones. There is nothing arbitrary about them. There is no other possible way to play with sound strategy and accurate mutual expectations. Nor would it be natural to call our regularity

a rule, except insofar as *any* regularity or any strategic maxim may sometimes be called a rule (and Shwayder certainly does not want to be that inclusive). This must be another kind of conformative behavior that Shwayder did not intend, for it is not—intuitively—action for the reason "That's the rule."

A last difference between Shwayder's rules and my conventions deserves to be mentioned only because the reader may have noticed it already. For Shwayder, a rule is a system of expectations likely to produce regular behavior. For me, a convention is a regularity in behavior produced by a system of expectations. Shwayder wants to allow (in one of his extensions from the fundamental case) for rules that are mostly broken; this generality may be plausible for his analysandum "rule," but not for my analysandum "convention." I am concerned with cases in which we have both the regularity and the expectations, and in such cases the regularity is the central thing. But the difference is superficial. It would vanish if we recast our analyses in the form, "A rule (convention) exists if and only if . . . "—not "A rule (convention) is . . ."

6. Imitation

Someone who is party to a convention conforms to a regularity because he has an interest in conforming if certain others do and because he believes—rightly—that they do. He acts as he does because he expects the others so to act. In short, he imitates them. But we should not conclude that any regularity which originates or persists by some sort of mutual imitation is therefore a convention. There are several cases in which each member of some population acts in conformity to some regularity because the others do and that regularity is nevertheless not a convention.

Sometimes people copy each other's actions—say, mannerisms—more or less unaware that they are doing so. Given a group composed entirely of such people, a mannerism can spread and persist by mutual imitation. But there is no preference involved on anyone's

part (unless we count every inclination as a transient preference). Each simply does something, not caring and scarcely knowing whether he does it or not. So *a fortiori* his action does not answer to any interest in so acting if the others do. The regularity produced is not a convention.

Sometimes people copy each other's preferences. A coffee drinker put among tea drinkers may somehow come to prefer tea. It is not that he prefers to drink coffee if the others drink coffee and tea if the others drink tea. At first he preferred coffee regardless of what the others preferred; later he prefers tea just as unconditionally. But it was his exposure to the tea drinkers that caused him to change. We can tell we are dealing with changeable unconditional preferences, not fixed conditional preferences, by observing a lag in his adaptation; or by observing his preferences regarding such things as a gamble that will entitle him to an unlimited supply of tea if all his neighbors switch to coffee. Given a group composed entirely of preference copiers, preferences and the actions answering to them could spread and persist by mutual imitation. But the regularity so produced would not be a convention, since the actions conforming to it would answer to an unconditional preference.

Sometimes people trust each other's practical judgments: crediting others with probably having good reasons for what they are doing, I may infer from their actions that they know something I do not—something that is a good reason for me to do the same. I may wear my raincoat because others do, thinking that probably they are wearing raincoats because they have heard a forecast of rain. This sort of imitation involves neither change of preferences nor preferences that are conditional on others' actions. Before and after I saw my neighbors in raincoats, my preference was to wear my raincoat if and only if it was going to rain. And that is my preference without regard to what others do. Their wearing of raincoats makes me wear one because I take their actions as evidence: evidence of their expectation of rain and therefore (indirectly, through my standing beliefs about the likely causes of their expectations) evidence

that it is likely to rain. A regularity might spread and persist in a group of people by just this sort of mutual imitation. It might happen one day that *everyone* in town wears his raincoat because he sees the others wearing theirs and infers—reasonably enough, perhaps,[15] but falsely—that they probably heard a forecast of rain. (The one who started it all must be an exception at first, since he put on his raincoat while the others were not wearing theirs. But whatever his original reason was, if he keeps his raincoat on later because he sees the others wearing theirs, he becomes just like the rest.) This regularity is not a convention; the preference that sustains it is not conditional on others' conforming.

Even when people do imitate each other because of their conditional preference for doing something if the others do, still the regularity that persists by this mutual imitation is not necessarily a convention. For the situation may not be one in which coincidence of interests predominates over conflict. Some sort of equilibrium is sustained, but it may not be a *coordination* equilibrium. In the example of pure conflict from the preceding section, for instance, each has a conditional preference for calling off if the other does. This preference permits a regularity whereby each calls off every time because the other does so. But we saw that this regularity is not a convention; although each is satisfied with his own choice (given his opponent's), neither is satisfied with his opponent's choice (given his own). Each wants to conform to the regularity of always calling off if the other also conforms to it; but each would like better to conform while the other does not. In the case of a genuine convention, on the other hand, each wants to conform if the others do, and each wants the others to conform if he does.

Now we have distinguished five pure species of regularity sustained by mutual imitation: convention itself and four counterfeits. They

[15] It may be that everyone was completely reasonable in inferring and acting as he did—although no one will think so when he learns what happened. The manifest irrationality of the group may not be due to any irrationality of its members. It is no mistake to expect rain when one sees people in raincoats, despite the bad results of doing so this time.

differ in the kind of imitation involved, the process whereby actions produce similar actions by others. Each case considered so far involves a single kind of imitation; but there could equally well be hybrids, regularities sustained by mutual imitation of several kinds mixed in varying proportions. It seems plausible that fashions, fads, panics, riots, and bandwagons as we know them are produced by several sorts of mutual imitation working together. Suppose we are wearing beige ties this year, each because the others do. Morgan follows the fashion without knowing he is doing so; when he picks a tie haphazardly, he just happens to pick the beige one. Jones is wearing beige ties because he likes the color; but, unknown to him, his tastes are caused by the prevailing fashion and will change with it. Griffith wears a beige tie because he falsely supposes the other beige wearers to have discovered some special functional virtue in beige which will benefit him too. Owen and Thomas both want to wear whatever color of tie the others will wear; but Owen hopes to find occasional nonconformists he can laugh at, whereas Thomas hopes there will be no nonconformists. Every man after his fashion follows the fashion, and the wearing of beige ties persists and spreads by mutual imitation of mixed kinds.

I do not count it a regularity by mutual imitation when a good idea catches on. Such a regularity spreads by imitation, as people see what others are doing and realize they would benefit by doing the same. But the imitation is not mutual—no two people learned the trick from each other—and the regularity does not persist by imitation. Once he starts, each goes on because he benefits by what he is doing whether other people go on doing it or not.