Foreword to an Extract from David K. Lewis’s “Convention”

Daniel B. Klein

Up to the time of David Hume’s *A Treatise of Human Nature* nearly all usage of the word *convention* meant the convening of persons to discuss or deliberate on a matter, or an agreement or resolution so arrived at—that is, a compact, contract, promise, or matter of consent (although sometimes “convention” was modified with a word like “tacit”). Hume himself acknowledged such meaning to be “the most usual sense of the word” (Hume 1998/1751, 98).

Hume’s *Treatise* advances another meaning, one that does not necessarily entail a convening nor a contract, compact, promise, or agreement. Hume emphasizes that feature. He says that a person may adopt a rule to abstain from another’s possessions from “a general sense of common interest” (Hume 2007/1740, 315). He says: “This convention is not of the nature of a promise: For even promises themselves…arise from human conventions” (ibid., 314). “In like manner are languages gradually establish’d by human conventions without any promise” (315). Language did not originate in the resolutions of convening syndics, since such syndics would presumably have to have had some language to conduct such deliberations.

That point about convening syndics is made by W. V. Quine (1969, xi) in his Foreword to David K. Lewis’s *Convention: A Philosophical Study*. Building on Hume and Thomas Schelling (1960), Lewis beautifully developed a definition of convention “along the lines of Hume” (1969, 3).

Here we republish a portion of Lewis’s book, which was his Ph.D. dissertation in philosophy at Harvard, originally published by Harvard University.

1. George Mason University, Fairfax, VA 22030.
Press in 1969. The portion republished here constitutes about 40 percent of the work. For permission to republish that portion, we are grateful to John M. Cooper of Princeton University, who represents the copyright holder. The present facsimile reproduction omits the original running headers and page numbers; pages 1–82 in the 1969 HUP edition correspond to pages 483–564 in the reproduction. (In this Foreword my citations to Lewis use HUP’s 1969 page numbers.)

**Lewisian convention**

Lewis defines coordination equilibrium. It is useful to contrast that with Nash equilibrium. Nash equilibrium can be described as the strategy profile that is a set of best responses. In the prisoners’ dilemma, the strategy profile (defect, defect) is a Nash equilibrium. Coordination equilibrium goes further: Not only does Jim like what he is doing given what everyone else is doing, but, further, he likes what each person is doing given what everyone else (including himself) is doing. This second aspect helps to ensure a certain “coincidence of interests” (Lewis 1969, 14). Thus, Jim not only finds his own choice to be a best response, he also finds the choice of each of the others agreeable, in the sense that he is glad of that choice, against the background of the remainder of the strategy profile. Lewis captures the two aspects of the definition as follows: “Let me define a coordination equilibrium as a [profile] in which no one would have been better off had any one agent alone acted otherwise, either himself or someone else” (ibid.). Every coordination equilibrium is a Nash equilibrium, but the converse is not true: (defect, defect) in the prisoners’ dilemma is not a coordination equilibrium because each player would prefer that the other choose ‘cooperate.’

All players find a coordination equilibrium to be agreeable in the sense just stated, but that does not mean everyone finds it particularly good or satisfactory. In fact, it may not even be Pareto efficient. Everyone driving on the left side of the road might be a coordination equilibrium. But if cars have the steering column on the left, then everyone driving on the right would make everyone better off. Thus, in the first case, when everyone drives on the left, each behaves agreeably, but the whole strategy profile, relative to one that has everyone driving on the right, is agreeable to no one, nor to an impartial spectator of the game. But the requirements of coordination equilibrium do not consider such multi-party changes to the strategy profile being tested.

Lewis adds the qualifier proper, to avoid oddities arising from ties in payoffs (1969, 22). He then defines coordination problems: “situations of interdependent decisions by two or more agents in which coincidence of interest predominates and in which there are two or more proper coordination equilibria” (ibid., 24).
Some people have taken multiplicity of Nash equilibria to be the defining feature of a coordination problem, but in the game called ‘Chicken,’ for example, there are two Nash equilibria \{(hawk, dove), (dove, hawk)\} and neither is a coordination equilibrium because in each the person playing dove would prefer that the other also play dove, so ‘Chicken’ is not a coordination problem. A coordination problem is represented by the ‘Battle of the Sexes’ game or the ‘Road’ game.

Lewis steps gracefully from formal game theory (coordination equilibria, etc.) to speaking in a looser way of a recurrent social situation, in the spirit of Schelling. As Edmund Burke (1990, 58) once put it, “A clear idea is…another name for a little idea.”

In defining convention, Lewis chaperons the reader through a progress of philosophizing. He works through a series of preliminary definitions (1969, 42, 58, 76). He notes problems of each, finds the need to define common knowledge, and then revises the definition, sometimes by loosening requirements, arriving at a fourth and final definition, given in two versions (ibid., 78–79).

In developing his definitions, Lewis speaks of a regularity in a recurrent situation. At first he explicitly makes the recurrent situation a coordination problem (1969, 42). But subsequently the form of the recurrent situation is loosened (ibid., 68–69) and quantifications are relaxed—for a few odd birds or malefactors, “children and the feeble-minded” (75), etc. In defining some regularity \(R\) to be a convention Lewis adds the presence of an alternative possible regularity \(R'\), to capture the idea that “there is no such thing as the only possible convention” (70). Built into the definition of convention, then, is a sense of a possible alternative agreeable way of behaving in the recurrent situation.

Regularity in a recurrent situation is one reason why a one-off matter of consent, the keeping of a particular promise or contract, would not, in that limited aspect, be convention: Fulfilling that particular contract is not a regularity in a recurrent situation. Lewis (1969, 84) explains another reason why even a promise pertaining to a recurrent situation might not qualify as a convention, namely that the promising itself might alter preferences in such a way as to make conforming to what was promised sufficiently unconditional so as to preclude the existence of (or even just common knowledge of the existence of) the sort of alternative regularity required for his definition of convention.

The relaxation of quantification and the loosening of form to a broad idea of a recurrent social situation have the result that the word “coordination” does not appear in Lewis’s final definition of convention. But the spirit of Lewis’s coordination formulations (coordination equilibrium, coordination problem) is very much

---

2. The second version of the final definition simply specifies indices \((d_0, d_1, d_2, d_3, d_4, d_5)\) to expresses degree, in lieu of the phrases “almost any instance,” “almost everyone conforms,” etc., in the first version.
preserved. Digesting those formal definitions is the way to ascend to Lewis’s final definition of convention (1969, 78).

**Why read Lewis on convention?**

Lewis’s work on convention is well known—as of 1 September 2019 the book had accumulated 7,342 Google Scholar citations—but it is our hope that the present republication will promote it further. Here is a list reasons, reflecting my pet interests, why you may benefit from the work:

- By following Lewis’s definition of convention, we may sharpen our thinking about distinctions between convention and other words used to approach culture, including *custom, moral, more, norm, practice, rule, regularity, institution, virtue, value,* and *culture.* Lewis’s work may improve our discrimination among such words.
- Lewis’s work can help us appreciate Hume’s thought, and to see its innovativeness.
- The matter of a possible alternative convention $R'$ helps to clarify the statement that commutative justice (in Adam Smith’s sense of the expression) is conventional: Within a context, the observed system (regularity) $R$ of commutative justice is a convention, for, within that same context, many an alternative ($R'$) would also be a possible convention. That among neighbors commutative justice is a necessary or natural convention, may be taken to mean that conformance to some such conventional system ($R, R'$, or some other) must exist. It is not commutative justice’s existence, as opposed to nonexistence, that is conventional, but rather the observed system of commutative justice, as opposed to some other system of commutative justice, notably a more primitive system now superseded, particularly by extensions of propertization. Commutative justice is a natural convention in the sense that necessarily, for any surviving community, a commutative-justice convention must exist among neighbors and other jural equals.
- Understanding the Humean tradition in terms of Lewis’s definition helps us to appreciate that tradition’s conventionalist theory of political authority—in particular to differentiate it from contractarian theories. This matter pertains to political theory generally, but also to the American identity. The rebel side of the War for Independence may have been imprinted with John Locke, *Cato’s Letters,* and Thomas Paine, but the Constitution and *Federalist Papers* bore more the spirit of men.
like Hume and Smith, who rejected social contract.
• Lewis, like Schelling, develops a body of ideas that might be placed under the rubric ‘mutual coordination.’ Under that rubric I would list the following as the major ideas: mutual coordination, coordination equilibrium, coordination game, coordination problem, focal point, common knowledge, convention, and emergent convention. I have addressed the polysemy of ‘coordination’ and have distinguished between the Schelling/Lewis meaning—‘mutual coordination’—and the meaning that until about 1970 had in economics been the much more prominent one, associated with Simon Newcomb, John Bates Clark, W. H. Hutt, Friedrich Hayek, Ronald Coase, Dennis Robertson, and many others, called ‘concatenate coordination’ (Klein 2012, 35–77). Studying Lewis puts one in a good position to see the distinction, the relationships between mutual coordination and concatenate coordination, and how they pertain to other ideas, such as cooperation.
• The present material nicely expositions the idea of common knowledge (Lewis 1969, 52–60).

What is the rest of the book about?

The present Lewis material might whet one’s appetite for the remainder of the book, so let me offer some remarks about that remaining 60 percent of the book. In the section that immediately follows the material reproduced here, Lewis serially contrasts convention with other concepts: “Agreement,” “Social Contracts,” “Norms,” “Rules,” “Conformative Behavior,” and “Imitation.” The series of contrasts not only instructs in how to think about each of the contrasted concepts, it also sharpens our understanding of convention, by highlighting facets of convention that differ from the contrasted concept.

In the final two sections of the book, Lewis turns to issues of the conventionality of language. He seeks to clarify issues of analyticity. At the end he writes: “I have given an account of the proper kind of analyticity—analyticity relative to a population of language users” (1969, 207). I am not confident in my understanding of these matters. Lewis affirms the idea of a (non-empty) category of analytic statements—analytic meaning that the statement is true in every possible world. Analytic statements stand in contrast to synthetic statements, which are true in some possible worlds and not in others. Thus, Lewis is sustaining a distinction between analytic and synthetic statements. Understand that statements are ‘within a language,’ and ‘a language’ can be, as it were, a rather sectarian language. The
sectarians may see to it that a certain statement is true in all possible worlds by
deciding that they will tailor their language so as to ensure its analyticity. I offer
some possible examples in a footnote.3 The decision to manage a language so as
to ensure analyticity might be perfectly reasonable. The final paragraph of Lewis
(1969) follows:

Analyticity is truth in all possible worlds. What is analytic for someone depends
jointly on the facts about the possible worlds and on the language he is using.
The language he is using depends on the conventions he is party to. And
these conventions are regularities in behavior, sustained by an interest in
coordination and an expectation that others will do their part. (Lewis 1969,
208)

I’m not sure what to make of analyticity and all that. The book discusses
“a convention of truthfulness” (1969, 152, 177ff.). One way to conform to a
convention of truthfulness, says Lewis, is “not to utter any [sentences] that are not
true” in the language your sect uses. As for being silent: “[H]old your tongue too
long—or at too high a cost to your social purposes—and you gradually turn from
a truthful user of [the language]…to a nonuser of [it]” (ibid., 178). Lewis also says
that a convention of truthfulness in one language implies alternative “regularities
of truthfulness in suitable alternative possible languages” (180). He distinguishes
between “truthfulness in a given possible language” and “truthfulness in the
language of one’s population, whichever possible language that may be.” He adds:
“It is the former, not the latter, which is conventional” (181). That point seems to
parallel the one I made above about commutative justice: The particular system of
commutative justice is conventional, but allegiance to one’s population’s system of
commutative justice, whichever that may be, is not conventional.

References


3. A sect might tailor its language so as to ensure the analyticity of the any of the following sentences: ‘The triangle has three sides.’ ‘The child was born of a mother.’ ‘The sum of the assets equals liabilities plus equity.’ ‘Y = C + I + G + NX.’ ‘The person maximized his utility.’ ‘If transaction costs were negligible and parties were aware of the relevant opportunities, those parties achieved an efficient outcome.’ ‘The moral sentiment relates to a sympathy.’


Convention: A Philosophical Study
Introduction, Chapter I, and Chapter II

David K. Lewis

TABLE OF CONTENTS

| Introduction | 483 | 1 |
| I | Coordination and Convention | 487 | 5 |
| 1. Sample Coordination Problems | 487 | 5 |
| 2. Analysis of Coordination Problems | 490 | 8 |
| 3. Solving Coordination Problems | 506 | 24 |
| 4. Convention | 518 | 36 |
| 5. Sample Conventions | 524 | 42 |
| II | Convention Refined | 534 | 52 |
| 1. Common Knowledge | 534 | 52 |
| 2. Knowledge of Conventions | 542 | 60 |
| 3. Alternatives to Conventions | 550 | 68 |
| 4. Degrees of Convention | 558 | 76 |
| 5. Consequences of Conventions | 562 | 80 |
| III | Convention Contrasted | 83 |
| 1. Agreement | 83 |
| 2. Social Contracts | 88 |
| 3. Norms | 97 |
| 4. Rules | 100 |
| 5. Conformative Behavior | 107 |
| 6. Imitation | 118 |
| IV | Convention and Communication | 122 |
| 1. Sample Signals | 122 |
| 2. Analysis of Signaling | 130 |
| 3. Verbal Signaling | 141 |
| 4. Conventional Meaning of Signals | 143 |
| 5. Meaning of Signals | 152 |
| V | Conventions of Language | 160 |
| 1. Possible Languages | 160 |
| 2. Grammars | 165 |
| 3. Semantics in a Possible Language | 173 |
| 4. Conventions of Truthfulness | 177 |
| 5. Semantics in a Population | 195 |
| Conclusion | 203 |
It is the profession of philosophers to question platitudes that others accept without thinking twice. A dangerous profession, since philosophers are more easily discredited than platitudes, but a useful one. For when a good philosopher challenges a platitude, it usually turns out that the platitude was essentially right; but the philosopher has noticed trouble that one who did not think twice could not have met. In the end the challenge is answered and the platitude survives, more often than not. But the philosopher has done the adherents of the platitude a service: he has made them think twice.

It is a platitude that language is ruled by convention. Words might be used to mean almost anything; and we who use them have made them mean what they do because somehow, gradually and informally, we have come to an understanding that this is what we shall use them to mean. We could perfectly well use these words otherwise—or use different words, as men in foreign countries do. We might change our conventions if we like.

To say only this is not to say much. It is not to portray language in the image of a calculus, precise and rigid. It is not to uphold “correct” speech against colloquial, or vice versa. It is not to say that all the languages we can think of are equally good, or that every feature of a serviceable language might just as well have been different. It is not to say that necessary truths are created by convention: only that necessary truths, like geological truths, are conventionally stated in these words rather than in those. It is not to exalt the powers of convention as some “conventionalist” philosophers do,
but only to insist that it is there. The platitude that there are conventions of language is no dogma of any school of philosophy, but commands the immediate assent of any thoughtful person—unless he is a philosopher.

For this mere platitude has been challenged. W. V. Quine questioned it in 1936 and later repudiated it outright. ¹ Morton White joined in the attack, ² and together they have persuaded some to share their doubts, and reduced many more to silence. Quine and White argue that the supposed conventions of language cannot be very much like the central, well-understood cases of convention. Conventions are agreements—but did we ever agree with one another to abide by stipulated rules in our use of language? We did not. If our ancestors did, how should that concern us, who have forgotten? In any case, the conventions of language could not possibly have originated by agreement, since some of them would have been needed to provide the rudimentary language in which the first agreement was made. We cannot even say what our conventions are, except by long trial and error. Did we know them better when we first adopted them? We have no concept of convention which permits language to be conventional; we are inclined to call some features of language conventional, but we cannot say why. We may indulge this inclination—Quine himself does ³—but we do not understand language any better for doing it. Conclusion: the conventions of language are a myth. The sober truth is that our use of language conforms to regularities—and that is all.


³At the end of “Carnap and Logical Truth” where he says: “The lore of our fathers... is a pale grey lore, black with fact and white with convention.”
We may protest, desperately, that there must be *something* to our notion of conventions of language, even if we cannot say what. When we are exposed to the notion we *do* all manage to get the idea, and all of us go on more or less alike in distinguishing between features of language we call conventional and features of language we do not. So we must mean something. Conventionality must at least be that, we know not what, which evokes a distinctive response in anyone who has been through our kind of education.

But how much better it would be to know what we are talking about: to have an analysis of convention in its full generality, including tacit convention not created by agreement. This book is my attempt at an analysis. I hope it is an analysis of our common, established concept of convention, so that you will recognize that it explains what you must have had in mind when you said that language—like many other activities—is governed by conventions. But perhaps it is not, for perhaps not all of us do share any one clear general concept of convention. At least, insofar as I had a concept of convention before I thought twice, this is either it or its legitimate heir. And what I call convention is an important phenomenon under any name. Language is only one among many activities governed by conventions that we did not create by agreeing and that we cannot describe.

My theory of convention had its source in the theory of games of pure coordination—a neglected branch of the general theory of games of von Neumann and Morgenstern, very different in method and content from their successful and better known theory of games of pure conflict. Coordination games have been studied by Thomas C. Schelling, and it is he who supplied me with the makings of an answer to Quine and White.

Yet, in the end, the theory of games is scaffolding. I can restate my analysis of convention without it. The result is a theory along the lines of Hume’s, in his discussion of the origin of justice and property. Convention turns out to be

a general sense of common interest; which sense all the members of the society express to one another, and which induces them to regulate their conduct by certain rules. I observe that it will be to my interest [e.g.] to leave another in the possession of his goods, provided he will act in the same manner with regard to me. When this common sense of interest is mutually expressed and is known to both, it produces a suitable resolution and behavior. And this may properly enough be called a convention or agreement betwixt us, though without the interposition of a promise; since the actions of each of us have a reference to those of the other, and are performed upon the supposition that something is to be performed on the other part.\footnote{A Treatise of Human Nature, III.ii.2.}
I | Coordination and Convention

1. Sample Coordination Problems

Use of language belongs to a class of situations with a conspicuous common character: situations I shall call coordination problems. I postpone a definition until we have seen a few examples. We begin with situations that might arise between two people—call them “you” and “I.”

1. Suppose you and I both want to meet each other. We will meet if and only if we go to the same place. It matters little to either of us where (within limits) he goes if he meets the other there; and it matters little to either of us where he goes if he fails to meet the other there. We must each choose where to go. The best place for me to go is the place where you will go, so I try to figure out where you will go and to go there myself. You do the same. Each chooses according to his expectation of the other’s choice. If either succeeds, so does the other; the outcome is one we both desired.

2. Suppose you and I are talking on the telephone and we are unexpectedly cut off after three minutes. We both want the connection restored immediately, which it will be if and only if one of us calls back while the other waits. It matters little to either of us whether he is the one to call back or the one to wait. We must each choose whether to call back, each according to his expectation of the other’s choice, in order to call back if and only if the other waits.

3. An example from Hume’s Treatise of Human Nature: Suppose you and I are rowing a boat together. If we row in rhythm, the boat goes smoothly forward; otherwise the boat goes slowly and erratically,
we waste effort, and we risk hitting things. We are always choosing whether to row faster or slower; it matters little to either of us at what rate we row, provided we row in rhythm. So each is constantly adjusting his rate to match the rate he expects the other to maintain.

Now we turn to situations among more than two people.

(4) Suppose several of us are driving on the same winding two-lane roads. It matters little to anyone whether he drives in the left or the right lane, provided the others do likewise. But if some drive in the left lane and some in the right, everyone is in danger of collision. So each must choose whether to drive in the left lane or in the right, according to his expectations about the others: to drive in the left lane if most or all of the others do, to drive in the right lane if most or all of the others do (and to drive where he pleases if the others are more or less equally divided).

(5) Suppose we are campers who have gone looking for firewood. It matters little to anyone in which direction he goes, but if any two go in the same direction they are likely to cover the same ground so that the one who gets there later finds no wood. Each must choose a direction to go according to his expectations about the others: one different from anyone else’s.

(6) Suppose several of us have been invited to a party. It matters little to anyone how he dresses. But he would be embarrassed if the others dressed alike and he dressed differently, since he knows that some discreditable explanation for that difference can be produced by whoever is so inclined. So each must dress according to his expectations about the way the others will dress: in a tuxedo if the others will wear tuxedos, in a clown suit if the others will wear clown suits (and in what he pleases if the others will dress in diverse ways).

(7) Suppose we are contented oligopolists. As the price of our raw material varies, we must each set new prices. It is to no one’s advantage to set his prices higher than the others set theirs, since if he does he tends to lose his share of the market. Nor is it to anyone’s advantage to set his prices lower than the others set theirs, since if he does he menaces his competitors and incurs their retaliation. So each must
set his prices within the range of prices he expects the others to set.

(8) An example from Rousseau’s *Discours sur l'inégalité*: Suppose we are in a wilderness without food. Separately we can catch rabbits and eat badly. Together we can catch stags and eat well. But if even one of us deserts the stag hunt to catch a rabbit, the stag will get away; so the other stag hunters will not eat unless they desert too. Each must choose whether to stay with the stag hunt or desert according to his expectations about the others, staying if and only if no one else will desert.

(9) Suppose we take it to be in our common interest that some scarce good, say grazing land, should be divided up somehow so that each of us can count on having the exclusive use of one portion. (Suppose nobody ever thinks it would be in his interest to help himself to someone else’s portion. The struggle, the harm to his neighbor, the bad example, the general loss of confidence, invariably seem to outweigh any gain.) It matters little to anyone who uses which portion, so long as people never try to use the same portion and no portion ever goes to waste. Each must choose which portion to use according to his expectations about the portions others will use and the portion they will leave for him.

(10) Suppose we are tradesmen. It matters little to any of us what commodities he takes in exchange for goods (other than commodities he himself can use). But if he takes what others refuse he is stuck with something useless, and if he refuses what others take he needlessly inconveniences his customers and himself. Each must choose what he will take according to his expectations about what he can spend—that is, about what the others will take: gold and silver if he can spend gold and silver, U.S. notes if he can spend U.S. notes, Canadian pennies if he can spend Canadian pennies, wampum if he can spend wampum, goats if he can spend goats, whatever may come along if he can spend whatever may come along, nothing if he can spend nothing.

(11) Suppose that with practice we could adopt any language in some wide range. It matters comparatively little to anyone (in the
long run) what language he adopts, so long as he and those around him adopt the same language and can communicate easily. Each must choose what language to adopt according to his expectations about his neighbors' language: English among English speakers, Welsh among Welsh speakers, Esperanto among Esperanto speakers, and so on.

2. Analysis of Coordination Problems

With these examples, let us see how to describe the common character of coordination problems.

Two or more agents must each choose one of several alternative actions. Often all the agents have the same set of alternative actions, but that is not necessary. The outcomes the agents want to produce or prevent are determined jointly by the actions of all the agents. So the outcome of any action an agent might choose depends on the actions of the other agents. That is why—as we have seen in every example—each must choose what to do according to his expectations about what the others will do.

Some combinations of the agents' chosen actions are equilibria: combinations in which each agent has done as well as he can given the actions of the other agents. In an equilibrium combination, no one agent could have produced an outcome more to his liking by acting differently, unless some of the others' actions also had been different. No one regrets his choice after he learns how the others chose. No one has lost through lack of foreknowledge.

This is not to say that an equilibrium combination must produce an outcome that is best for even one of the agents (though if there is a combination that is best for everyone, that combination must be an equilibrium). In an equilibrium, it is entirely possible that some or all of the agents would have been better off if some or all had acted differently. What is not possible is that any one of the agents would have been better off if he alone had acted differently and all the rest had acted just as they did.
We can illustrate equilibria by drawing *payoff matrices* for coordination problems between two agents. Call the agents *Row-chooser* and *Column-chooser*. We represent Row-chooser's alternative actions by labeled rows of the matrix, and Column-chooser's by labeled columns. The squares then represent combinations of the agents' actions and the expected outcomes thereof. Squares are labeled with two *payoffs*, numbers somehow measuring the desirability of the expected outcome for Row-chooser and Column-chooser.\(^1\) Row-chooser's payoff is at the lower left, Column-chooser's at the upper right.

Thus the matrix of Figure 1 might represent a simple version of example (1), where \(R_1, R_2,\) and \(R_3\) are Row-chooser's actions of \(C_1, C_2,\) and \(C_3\) are Column-chooser's actions of going to places \(P_1, P_2,\) and \(P_3\) respectively, and \(C_1, C_2,\) and \(C_3\) are Column-chooser's actions of going to places \(P_1, P_2,\) and \(P_3\) respectively. The equilibria are the three combinations in which Row-

\[\begin{array}{ccc}
C1 & C2 & C3 \\
R1 & 1 & 0 & 0 \\
& meet & 0 & 0 \\
R2 & 0 & 1 & 0 \\
& meet & 1 & 0 \\
R3 & 0 & 0 & 1 \\
& & & meet \\
\end{array}\]

Figure 1

\(^1\)My account will demand no great sophistication about these numerical measures of desirability. If a foundation is required, it could be provided by decision theory as developed, for instance, by Richard Jeffrey in *The Logic of Decision* (New York: McGraw-Hill, 1965). I take it that decision theory applies in some approximate way to ordinary rational agents with imperfectly coherent preferences; our payoffs need never be more than rough indications of strength of preference.
chooser and Column-chooser go to the same place and meet there: \( \langle R1, C1 \rangle \), \( \langle R2, C2 \rangle \), and \( \langle R3, C3 \rangle \). For instance, \( \langle R2, C2 \rangle \) is an equilibrium by definition because Row-chooser prefers it to \( \langle R1, C2 \rangle \) or \( \langle R3, C2 \rangle \), and Column-chooser prefers it to \( \langle R2, C1 \rangle \) or \( \langle R2, C3 \rangle \). Both are indifferent between the three equilibria.

But suppose we change the example so that Row-chooser and Column-chooser care where they go, though not nearly so much as they care whether they meet. The new payoff matrix might be as shown in Figure 2. The equilibria remain the same: \( \langle R1, C1 \rangle \), \( \langle R2, C2 \rangle \), and \( \langle R3, C3 \rangle \). But Row-chooser and Column-chooser are no longer indifferent between the equilibria. \( \langle R1, C1 \rangle \) is the best possible outcome for both; \( \langle R3, C3 \rangle \) is the worst equilibrium outcome for both, though both prefer it to the nonequilibrium outcomes. Or if the payoff matrix were as shown in Figure 3, then \( \langle R1, C1 \rangle \) would be Row-chooser’s best outcome and Column-chooser’s worst equilibrium outcome; \( \langle R3, C3 \rangle \) would be Column-chooser’s best outcome and Row-chooser’s worst equilibrium outcome. No outcome would be best for both.

There seems to be a difference between equilibrium combinations in which every agent does the same action and equilibrium combinations in which agents do different actions. This difference is spurious, however. We say that the agents do the same action if they do actions
of the same kind, particular actions falling under a common descrip­
tion. But actions can be described in any number of ways, of which
none has any compelling claim to primacy. For any combination
of actions, and a fortiori for any equilibrium combination of actions,
there is some way of describing the agents' alternative actions so
that exactly those alternative actions in the given combination fall
under a common description. Any combination, equilibrium or not,
is a combination of actions of a same kind (a kind that excludes
all the agents' alternative actions). Whether it can be called a combi­
nation in which every agent does the same action depends merely
on the naturalness of that classification.

Consider example (2). If we have in mind these action-descriptions,

\[ R_1 \text{ or } C_1: \text{ calling back} \]
\[ R_2 \text{ or } C_2: \text{ not calling back} \]

we draw the payoff matrix shown in Figure 4 and think of the case
as one in which the equilibria \( \langle R_1, C_2 \rangle \) and \( \langle R_2, C_1 \rangle \) are combina­
tions in which the agents do different actions. But if we have in mind
these action-descriptions,

\[ R'_1 \text{ or } C'_1: \text{ calling back if and only if one is the original caller} \]
\[ R'_2 \text{ or } C'_2: \text{ calling back if and only if one is not the original caller} \]
we draw the payoff matrix shown in Figure 5 and think of the case as one in which the equilibria \(\langle R1', C1' \rangle\) and \(\langle R2', C2' \rangle\) are combinations in which the agents do the same action. But what makes the first pair of action-descriptions more natural than the second? And so what if it is?

We might say that coordination problems are situations in which several agents try to achieve uniformity of action by each doing whatever the others will do. But this is a dangerous thing to say, since it is true of a coordination problem only under suitable descriptions of actions, and sometimes the descriptions that make it true would strike us as contrived—so, for instance, in examples (2), (5), (9), and perhaps (4). What is important about the uniform combinations we are interested in is not that they are—under some description—uniform, but that they are equilibria.

Of course this is not to say that coordination problems are distin-
guished by the presence of equilibria. Indeed the bulk of the mathematical theory of games is precisely the theory of equilibrium combinations (known also as saddle points or solutions) in situations of the opposite kind: pure conflict of interest between two agents, as in Figure 6.

Figure 6

<table>
<thead>
<tr>
<th></th>
<th>C1</th>
<th>C2</th>
<th>C3</th>
</tr>
</thead>
<tbody>
<tr>
<td>R1</td>
<td>0</td>
<td>-.5</td>
<td>-.5</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>.5</td>
<td>.5</td>
</tr>
<tr>
<td>R2</td>
<td>.5</td>
<td>1</td>
<td>-1</td>
</tr>
<tr>
<td></td>
<td>-.5</td>
<td>-1</td>
<td>1</td>
</tr>
<tr>
<td>R3</td>
<td>.5</td>
<td>-1</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>-.5</td>
<td>1</td>
<td>-1</td>
</tr>
</tbody>
</table>

In general, pure conflict can be represented by a payoff matrix in which the agents’ payoffs (perhaps after suitable linear rescaling) sum to zero in every square. This is to say that one agent’s losses are the others’ gains, and vice versa. Yet there are equilibria in pure conflict. In the example shown, (R₁, C₁) is an equilibrium: Row-chooser prefers it to (R₂, C₁) or (R₃, C₁), and Column-chooser prefers it to (R₁, C₁) or (R₁, C₃).

Schelling argues for a “reorientation of game theory” in which games—problems of interdependent decision—are taken to range over a spectrum with games of pure conflict and games of pure coordination as opposite limits. Games of pure conflict, in which the

---

²There is no point in changing the definition to let the sum be a constant other than zero. By allowing rescaling, we already have full generality. Without rescaling, we would not reach full generality just by allowing nonzero constant sums. And by allowing linear rescaling, we make clear why—despite appearance—our definitions do not depend on any problematic interpersonal comparison of desirabilities.

agents' interests are perfectly opposed, can be defined as we have just seen. *Games of pure coordination*, in which the agents’ interests coincide perfectly, are games in which the agents’ payoffs (perhaps after suitable linear rescaling) are equal in every square. Other games are mixtures in varying proportions of conflict and coordination, of opposition and coincidence of interests.

My coordination problems such as (1)–(11) are among the situations at or near the pure coordination end of Schelling’s spectrum. I do not want to require perfect coincidence of interests. For instance, I allowed imperfect coincidence of interests in those versions of example (1) in which Row-chooser and Column-chooser care somewhat where they go, though much less than they care whether they meet. We recall the payoff matrices of Figures 2 and 3 (pp. 10–11). In several squares, the payoffs are not quite equal. No linear rescaling of either matrix could make them equal in every square at once.

I want, however, to confine my attention to situations in which coincidence of interest predominates: that is, in which the differences between different agents’ payoffs in any one square (perhaps after suitable linear rescaling) are small compared to some of the differences between payoffs in different squares. So they are in the matrices of Figures 2 and 3; the largest difference within one square is .5, whereas the largest difference between payoffs in different squares is 1.5.

An equilibrium, we recall, is a combination in which no one would have been better off had he alone acted otherwise. Let me define a *coordination equilibrium* as a combination in which no one would have been better off had any one agent alone acted otherwise, either himself or someone else. Coordination equilibria are equilibria, by the definitions. Equilibria in games of pure coordination are always coordination equilibria, since the agents’ interests coincide perfectly. Any game of pure coordination has at least one coordination equilibrium, since it has at least one outcome that is best for all. But coordination equilibria are by no means confined to games of pure coordination. They are common in situations with mixed opposition.
and coincidence of interests. They can occur even in games of pure conflict: \( \langle R1, C1 \rangle \) in Figure 7 is a coordination equilibrium.

Most versions of our sample coordination problems are not games of pure coordination; but they all have coordination equilibria. We have noticed that the versions of the meeting-place problem shown in Figures 2 and 3 are not games of pure coordination; but their equilibria—\( \langle R1, C1 \rangle \), \( \langle R2, C2 \rangle \), and \( \langle R3, C3 \rangle \) in both versions—are coordination equilibria.

This is not to say that all the equilibria in a coordination problem must be coordination equilibria. Take still another version of example (1). Suppose there is a fourth place, \( P4 \). Row Chooser and Column Chooser both like to go to \( P4 \) alone, but a meeting at \( P4 \) would detract from their enjoyment of going to \( P4 \) and \( P4 \) would be of little use as a meeting place. So we have the matrix shown in Figure 8, with the usual coordination equilibria \( \langle R1, C1 \rangle \), \( \langle R2, C2 \rangle \), \( \langle R3, C3 \rangle \) and a new noncoordination equilibrium \( \langle R4, C4 \rangle \). It is an equilibrium because Row Chooser prefers it to \( \langle R1, C4 \rangle \), \( \langle R2, C4 \rangle \), or \( \langle R3, C4 \rangle \), and Column Chooser prefers it to \( \langle R4, C1 \rangle \), \( \langle R4, C2 \rangle \), or \( \langle R4, C3 \rangle \). It is not a coordination equilibrium because not all—in fact, none—of these preferences are shared by Row Chooser and Column Chooser. Yet this version of (1) does not seem significantly different from the others. The situation still has that distinctive character which I introduced by means of my eleven examples. So let us tolerate noncoordination equilibria in coordination problems.
All my sample coordination problems have two or more different coordination equilibria. This multiplicity is important to the distinctive character of coordination problems and ought to be included in their definition. If there is no considerable conflict of interest, the task of reaching a unique coordination equilibrium is more or less trivial. It will be reached if the nature of the situation is clear enough so that everybody makes the best choice given his expectations, everybody expects everybody else to make the best choice given his expectations, and so on. These conditions do not ensure coordination if there are multiple coordination equilibria, as we shall see.

Many of the situations with unique coordination equilibria are still
more trivial (and more deserving of exclusion). For instance, any situation in which all the agents have [strictly] dominant choices—actions they prefer no matter what the others do—can have only one equilibrium (and a fortiori only one coordination equilibrium), namely, the combination of dominant choices. A combination of dominant choices must be an equilibrium; but it might not be a coordination equilibrium, as in the well-known Prisoner’s Dilemma, shown in Figure 9, in which $R_I$ and $C_I$ (treacherous confession, in the usual story) are dominant and their combination $\langle R_I, C_I \rangle$ is a noncoordination equilibrium.

We might guess that there is dominance in any game of pure coordination with a unique equilibrium: that all, or at least some, agents have dominant, or at least dominated, choices. (A [strictly] dominated choice is one such that, no matter how the others choose, you could have made some other choice that would have been better. If one choice is dominant, another must be dominated; but not vice versa, since which other choice would have been better for you may depend on how the others chose.) There is this much truth in the guess: in any finite two-person game of pure coordination with a unique equilibrium, at least one action of one of the agents is dominated. Proof:

Let $P(\langle R_j, C_k \rangle)$ represent the payoff at the combination $\langle R_j, C_k \rangle$, equal for Row-chooser and Column-chooser.

Take a suitable game with $m$ rows and $n$ columns. Assume without loss of generality that its rows and columns are so arranged that for any combination $\langle R_i, C_i \rangle$ on the diagonal and any combination $\langle R_j, C_k \rangle$ such that $j \geq i$ and $k \geq i$, $P(\langle R_j, C_k \rangle) \leq P(\langle R_i, C_i \rangle)$. In particular, $\langle R_I, C_I \rangle$ must be the unique equilibrium, and $P(\langle R_I, C_I \rangle)$ must exceed every other payoff in the game.

If $\langle R_I, C_I \rangle$ is the only diagonal combination that is either a row-maximum or a column-maximum, then $R_m$ (if $m \geq n$) or $C_n$ (if $n \geq m$) must be dominated.
Otherwise let \( \langle Ra, Ca \rangle \), \( a \neq 1 \), be the rightmost diagonal combination which is either a row-maximum or a column-maximum. It is not both, since it is not an equilibrium. Suppose without loss of generality that it is a row-maximum.

Unless \( Ra \) is strictly dominated, there is a column-maximum on \( Ra \); let \( \langle Ra, Cb \rangle \) be the rightmost one. \( \langle Ra, Cb \rangle \) is not a row-maximum since it is not an equilibrium, so \( P(\langle Ra, Ca \rangle) > P(\langle Ra, Cb \rangle) \).

Unless \( Cb \) is strictly dominated, there is a row-maximum on \( Cb \); let \( \langle Ra', Cb \rangle \) be the lowest one. Since \( \langle Ra, Cb \rangle \) is a column-maximum, \( P(\langle Ra, Cb \rangle) \geq P(\langle Ra', Cb \rangle) \), so \( P(\langle Ra, Ca \rangle) > P(\langle Ra', Cb \rangle) \).

Unless \( Ra' \) is strictly dominated there is a column-maximum on \( Ra' \); let \( \langle Ra', Cb' \rangle \) be the rightmost one; \( P(\langle Ra, Ca \rangle) > P(\langle Ra', Cb' \rangle) \).

Unless \( Cb' \) is strictly dominated, there is a row-maximum on \( Cb' \); let \( \langle Ra'', Cb' \rangle \) be the lowest one; \( P(\langle Ra, Ca \rangle) > P(\langle Ra'', Cb' \rangle) \).

Unless \( Ra'' \) is strictly dominated, there is a column-maximum on \( Ra'' \); let \( \langle Ra'', Cb'' \rangle \) be the rightmost one; \( P(\langle Ra, Ca \rangle) > P(\langle Ra'', Cb'' \rangle) \). And so on.

If \( \langle Rj, Ci \rangle \) is a column-maximum and \( P(\langle Ra, Ca \rangle) > P(\langle Rj, Ci \rangle) \), then \( \langle Rj, Ci \rangle \) is above the diagonal. For otherwise \( j \geq i \), so \( P(\langle Rj, Ci \rangle) \leq P(\langle Ri, Ci \rangle) \). And since \( \langle Rj, Ci \rangle \) is a column-maximum, \( P(\langle Rj, Ci \rangle) = P(\langle Ri, Ci \rangle) \). Then \( \langle Ri, Ci \rangle \) is also a column-maximum, and it is to the right of \( \langle Ra, Ca \rangle \) since \( P(\langle Ra, Ca \rangle) > P(\langle Ri, Ci \rangle) \). But that is contrary to our choice of \( \langle Ra, Ca \rangle \).

In particular: \( \langle Ra, Cb \rangle, \langle Ra', Cb' \rangle, \langle Ra'', Cb'' \rangle \), etc. are above the diagonal.

By a parallel argument, if \( \langle Rj, Ci \rangle \) is a row-maximum and \( P(\langle Ra, Ca \rangle) > P(\langle Rj, Ci \rangle) \), then \( \langle Rj, Ci \rangle \) is below the diagonal. In particular: \( \langle Ra', Cb \rangle, \langle Ra'', Cb' \rangle \), etc. are below the diagonal.

Therefore the sequence of combinations we were constructing
moves back and forth across the diagonal, as shown in Figure 10, so that $a < a' < a'' \ldots$ and $b < b' < b'' \ldots$. Since the game is finite, these sequences terminate, which can happen only if one of $Ra$, $Cb$, $Ra'$, $Cb'$, $Ra''$, $Cb''$ etc. is strictly dominated.

Figure 10

The deletion of a dominated action in a finite two-person game of pure coordination with a unique equilibrium leaves a new game, which is itself a finite two-person game of pure coordination with a unique equilibrium. So the deletion can be repeated. By successive deletions of dominated actions, the game is transformed into a situation that is patently trivial because Row-chooser and Column-chooser each have only one available action. The outcome is determined by the fact that everybody ignores dominated actions, everybody expects everybody else to ignore dominated actions, and so on.

The result just proved cannot, unfortunately, be strengthened in any of the ways one might hope. It does not carry over to infinite
two-person games; Figure 11 is a counterexample. It does not carry over to finite three-person games; Figure 12 is a counterexample.

<table>
<thead>
<tr>
<th></th>
<th>C1</th>
<th>C2</th>
<th>C3</th>
<th>C4</th>
<th>C5</th>
<th>...</th>
</tr>
</thead>
<tbody>
<tr>
<td>R1</td>
<td>64</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>64</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>R2</td>
<td>32</td>
<td>16</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>32</td>
<td>16</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>R3</td>
<td>0</td>
<td>8</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>8</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>R4</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>R5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>.5</td>
<td>.25</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>.5</td>
<td>.25</td>
<td></td>
</tr>
</tbody>
</table>

Figure 11

(Call the third agent's choices levels $L_1$ and $L_2$; write his payoffs in the centers of the squares.) It cannot be strengthened for the finite two-person case; Figure 13 is an example with no dominant action and only a single dominated action (and that one is dominated only
by all the alternatives together). Therefore we cannot say that dominance is responsible for all cases of unique equilibria in games of pure coordination.

To exclude trivial cases, a coordination problem must have more than one coordination equilibrium. But that requirement is not quite strong enough. Figure 14 shows two matrices in which, sure enough,

there are multiple coordination equilibria (two on the left, four on the right). Yet there is still no need for either agent to base his choice on his expectation about the other’s choice. There is no need for them to try for the same equilibrium—no need for coordination—since if
they try for different equilibria, some equilibrium will nevertheless be reached. These cases exhibit another kind of triviality, akin to the triviality of a case with a unique coordination equilibrium.

A combination is an equilibrium if each agent likes it at least as well as any other combination he could have reached, given the others' choices. Let us call it a proper equilibrium if each agent likes it better than any other combination he could have reached, given the others' choices. In a two-person matrix, for instance, a proper equilibrium is preferred by Row-chooser to all other combinations in its column, and by Column-chooser to all other combinations in its row. In the matrices in Figure 14, there are multiple coordination equilibria, but all of them are improper.

There is no need to stipulate that all equilibria in a coordination problem must be proper; it seems that the matrix in Figure 15 ought to be counted as essentially similar to our clear examples of coordination problems, despite the impropriety of its equilibrium \( \langle R3, C3 \rangle \). The two proper coordination equilibria—\( \langle R1, C1 \rangle \) and \( \langle R2, C2 \rangle \)—are sufficient to keep the problem nontrivial. I stipulate instead that a coordination problem must contain at least two proper coordination equilibria.

This is only one—the strongest—of several defendable restrictions. We might prefer a weaker restriction that would not rule out matrices like those in Figure 16. But a satisfactory restriction would be com-
plicated and would entail too many qualifications later. And situa-
tions like those of Figure 16 can be rescued even under the strong
restriction we have adopted. Let $R_2'$ be the disjunction of $R_2$ and
$R_3$, and $C_1'$ the disjunction of $C_2$ and $C_3$ in the left-hand matrix.
Then the same situation can be represented by the new matrix in
Figure 17, which does have two proper coordination equilibria. The
right-hand matrix can be consolidated in a similar way. But matrices
like the one in Figure 18, which are ruled out by the strong restric-
tion, and ought to be ruled out, cannot be rescued by any such
consolidation.
To sum up: Coordination problems—situations that resemble my eleven examples in the important respects in which they resemble one another\(^4\)—are situations of interdependent decision by two or more agents in which coincidence of interest predominates and in which there are two or more proper coordination equilibria. We could also say—though less informatively than one might think—that they are situations in which, relative to some classification of actions, the agents have a common interest in all doing the same one of several alternative actions.

3. Solving Coordination Problems

Agents confronted by a coordination problem may or may not succeed in each acting so that they reach one of the possible coordination equilibria. They might succeed just by luck, although some of them choose without regard to the others' expected actions (doing so perhaps because they cannot guess what the others will do, perhaps because the chance of coordination seems so small as to be negligible).

\(^4\)See Michael Slote, "The Theory of Important Criteria," *Journal of Philosophy*, 63 (1966), pp. 211–224. Slote shows that we commonly introduce a class by means of examples and take the defining features of the class to be those distinctive features of our examples which seem important for an understanding of their character. That is what I take myself to be doing here and elsewhere.
But they are more likely to succeed—if they do—through the agency of a system of suitably concordant mutual expectations. Thus in example (1) I may go to a certain place because I expect you to go there, while you go there because you expect me to; in example (2) I may call back because I expect you not to, while you do not because you expect me to; in example (4) each of us may drive on the right because he expects the rest to do so; and so on. In general, each may do his part of one of the possible coordination equilibria because he expects the others to do theirs, thereby reaching that equilibrium.

If an agent were completely confident in his expectation that the others would do their parts of a certain proper coordination equilibrium, he would have a decisive reason to do his own part. But if—as in any real case—his confidence is less than complete, he must balance his preference for doing his part if the others do theirs against his preferences for acting otherwise if they do not. He has a decisive reason to do his own part if he is sufficiently confident in his expectation that the others will do theirs. The degree of confidence which is sufficient depends on all his payoffs and sometimes on the comparative probabilities he assigns to the different ways the others might not all do their parts, in case not all of them do. For instance, in the coordination problem shown in Figure 19, Row-chooser should do his part of the coordination equilibrium \((R_1, C_1)\) by choosing \(R_1\) if he has more than .5 confidence that Column-chooser will do his part by choosing \(C_1\). But in the coordination problems shown in Figure 20, Row-chooser should choose \(R_1\) only if he has more...
than .9 confidence that Column-chooser will choose $C_1$. If he has, say, .8 confidence that Column-chooser will choose $C_1$, he would do better to choose $R_2$, sacrificing his chance to achieve coordination at $(R_1, C_1)$ in order to hedge against the possibility that his expectation was wrong. And in the coordination problem shown in Figure 21, Row-chooser might be sure that if Column-chooser fails to do

his part of $(R_1, C_1)$, at least he will choose $C_2$, not $C_3$; if so, Row-chooser should choose $R_1$ if he has more than .5 confidence that Column-chooser will choose $C_1$. Or Row-chooser might think that if Column-chooser fails to choose $R_1$, he is just as likely to choose $C_3$ as to choose $C_2$; if so, Row-chooser should choose $R_1$ only if he has more than .9 confidence that Column-chooser will choose $C_1$. Or Row-chooser might be sure that if Column-chooser does not choose $C_1$, he will choose $C_3$ instead; if so, Row-chooser's minimum sufficient degree of confidence is about .95. The strength of concordant expectation needed to produce coordination at a certain equilibrium is a measure of the difficulty of achieving coordination there,
since however the concordant expectations are produced, weaker expectations will be produced more easily than stronger ones. (We can imagine cases in which so much mutual confidence is required to achieve coordination at an equilibrium that success is impossible. Imagine that a millionaire offers to distribute his fortune equally among a thousand men if each sends him $10; if even one does not, the millionaire will keep whatever he is sent. I take it that no matter what the thousand do to increase their mutual confidence, it is a practical certainty that the millionaire will not have to pay up. So if I am one of the thousand, I will keep my $10.)

We may achieve coordination by acting on our concordant expectations about each other's actions. And we may acquire those expectations, or correct or corroborate whatever expectations we already have, by putting ourselves in the other fellow's shoes, to the best of our ability. If I know what you believe about the matters of fact that determine the likely effects of your alternative actions, and if I know your preferences among possible outcomes and I know that you possess a modicum of practical rationality, then I can replicate your practical reasoning to figure out what you will probably do, so that I can act appropriately.

In the case of a coordination problem, or any other problem of interdependent decision, one of the matters of fact that goes into determining the likely effects of your alternative actions is my own action. In order to figure out what you will do by replicating your practical reasoning, I need to figure out what you expect me to do.

I know that, just as I am trying to figure out what you will do by replicating your reasoning, so you may be trying to figure out what I will do by replicating my reasoning. This, like anything else you might do to figure out what I will do, is itself part of your reasoning. So to replicate your reasoning, I may have to replicate your attempt to replicate my reasoning.

This is not the end. I may reasonably expect you to realize that, unless I already know what you expect me to do, I may have to try to replicate your attempt to replicate my reasoning. So I may expect you to try to replicate my attempt to replicate your attempt to
replicate my reasoning. So my own reasoning may have to include an attempt to replicate your attempt to replicate my attempt to replicate your attempt to replicate my reasoning. And so on.

Before things get out of hand, it will prove useful to introduce the concept of higher-order expectations, defined by recursion thus:

A first-order expectation about something is an ordinary expectation about it.

An \((n + 1)\)th-order expectation about something \((n \geq 1)\) is an ordinary expectation about someone else's \(n\)th-order expectation about it.

For instance, if I expect you to expect that it will thunder, then I have a second-order expectation that it will thunder.

Whenever I replicate a piece of your practical reasoning, my second-order expectations about matters of fact, together with my first-order expectations about your preferences and your rationality, justify me in forming a first-order expectation about your action. In the case of problems of interdependent decision—for instance, coordination problems—some of the requisite second-order expectations must be about my own action.

Consider our first sample coordination problem: a situation in which you and I want to meet by going to the same place. Suppose that after deliberation I decide to come to a certain place. The fundamental practical reasoning which leads me to that choice is shown in Figure 22. (In all diagrams of this kind, heavy arrows represent implications; light arrows represent causal connections between the mental states or actions of a rational agent.) And if my premise for this reasoning—my expectation that you will go there—was obtained by replicating your reasoning, my replication is shown in Figure 23. And if my premise for this replication—my expectation that you will expect me to go there—was obtained by replicating your replication of my reasoning, my replication of your replication is shown in Figure 24. And so on. The whole of my reasoning (simplified by disregarding the rationality premises) may be represented as in
I desire that I go there on condition that you will go there

I expect that you will go there

I have reason to desire that I go there

I go there

Figure 22

I expect that you desire that you go there on condition that I will go there

I expect that you expect that I will go there

I have reason to expect that you have reason to desire that you go there

I have reason to expect that you will go there

I expect that you will go there

Figure 23

I expect that you are rational to a certain degree
Figure 24 for whatever finite number of stages it may take for me to use whatever higher-order expectations may be available to me regarding our actions and our conditional preferences. Replications are nested to some finite depth: my reasoning (outer boundary) con-
I expect that you expect that I desire that I go there on condition that you will go there

I expect that you desire that you go there on condition that I will go there

I desire that I go there on condition that you will go there

I go there

Figure 25

contains a replication of yours (next boundary), which contains a replication of your replication of mine (next boundary), and so on.

So if I somehow happen to have an nth-order expectation about action in this two-person coordination problem, I may work outward through the nested replications to lower- and lower-order expectations about action. Provided I go on long enough, and provided all the needed higher-order expectations about preferences and rationality are available, I eventually come out with a first-order expectation about your action—which is what I need in order to know how I should act.

Clearly a similar process of replication is possible in coordination problems among more than two agents. In general, my higher-order
expectations about something are my expectations about $x_1$'s expectations about $x_2$'s expectations . . . about it. (The sequence $x_1, x_2$ . . . may repeat, but $x_1$ cannot be myself and no one can occur twice in immediate succession.) So when $m$ agents are involved, I can have as many as $(m - 1)^n$ different $n$th-order expectations about anything, corresponding to the $(m - 1)^n$ different admissible sequences of length $n$. Replication in general is ramified: it is built from stages in which $m - 1$ of my various $(n + 1)$th-order expectations about action, plus ancillary premises, yield one of my $n$th-order expectations about action. I suppressed the ramification by setting $m = 2$, but the general case is the same in principle.

Note that replication is not an interaction back and forth between people. It is a process in which one person works out the consequences of his beliefs about the world—a world he believes to include other people who are working out the consequences of their beliefs, including their belief in other people who . . . By our interaction in the world we acquire various high-order expectations that can serve us as premises. In our subsequent reasoning we are windowless monads doing our best to mirror each other, mirror each other mirroring each other, and so on.

Of course I do not imagine that anyone will solve a coordination problem by first acquiring a seventeenth-order expectation from somewhere and then sitting down to do his replications. For one thing, we rarely do have expectations of higher order than, say, fourth. For another thing, any ordinary situation that could justify a high-order expectation would also justify low-order expectations directly, without recourse to nested replications.

All the same, given the needed ancillary premises, an expectation of arbitrarily high order about action does give an agent one good reason for a choice of action. The one may, and normally will, be one reason among the many which jointly suffice to justify his choice. Suppose the agent is originally justified somehow in having expectations of several orders about his own and his partners' actions. And suppose the ancillary premises are available. Then each of his original expectations independently gives him a reason to act one way or
another. If he is lucky, all these independent reasons will be reasons for the same action. Then that action is strongly, because redundantly, justified; he has more reason to do it than could have been provided by any one of his original expectations by itself.

I said earlier that coordination might be rationally achieved with the aid of concordant mutual expectations about action. We have seen that these may be derived from first- and higher-order expectations about action, preferences, and rationality. So we generalize: coordination may be rationally achieved with the aid of a system of concordant mutual expectations, of first or higher orders, about the agents' actions, preferences, and rationality.

The more orders of expectation about action contribute to an agent's decision, the more independent justifications the agent will have; and insofar as he is aware of those justifications, the more firmly his choice will be determined. Circumstances that will help to solve a coordination problem, therefore, are circumstances in which the agents become justified in forming mutual expectations belonging to a concordant system. And the more orders, the better.

In considering how to solve coordination problems, I have postponed the answer that first comes to mind: by agreement. If the agents can communicate (without excessive cost), they can ensure a common understanding of their problem by discussing it. They can choose a coordination equilibrium—an arbitrary one, or one especially good for some or all of them, or one they can reach without too much

---

5 Michael Scriven, in “An Essential Unpredictability in Human Behavior,” *Scientific Psychology: Principles and Approaches*, ed. B. B. Wolman (New York: Basic Books, 1965), has discussed mutual replication of practical reasoning between agents in a game of conflict who want *not* to conform to each other's expectations. There is a cyclic alternation: from my \( (n + 4) \)th-order expectation that I will go to Minsk to my \( (n + 3) \)th-order expectation that you will go to Pinsk to my \( (n + 2) \)th-order expectation that I will go to Pinsk to my \( (n + 1) \)th-order expectation that you will go to Minsk to my \( n \)th-order expectation that I will go to Pinsk . . . Scriven notices that we cannot both act on complete and accurate replications of each other's reasoning. He takes this to prove human unpredictability. But perhaps it simply proves that the agents cannot both have enough time to finish their replications, since the time either needs increases with the time the other uses. See David Lewis and Jane Richardson, "Scriven on Human Unpredictability," *Philosophical Studies*, 17 (1966), pp. 69–74.
mutual confidence. And each can assure the rest that he will do his part of the chosen equilibrium. Coordination by means of an agreement is not, of course, an alternative to coordination by means of concordant mutual expectations. Rather, agreement is one means of producing those expectations. It is an especially effective means, since it produces strong concordant expectations of several orders.

Suppose you and I want to meet tomorrow; today we happen to meet, and we make an appointment. Each thereby gives evidence of his interest in going where the other goes and of his intention to go to a certain place. By observing this evidence, we form concordant first-order expectations about each other's preferences and action. By observing each other observing it, we may also form concordant second-order expectations. By observing each other observing each other observing it, we may even form concordant third-order expectations. And so on; not forever, of course, but limited by the amount of reasoning we do and the amount we ascribe to each other—perhaps one or two steps more. The result is a system of concordant mutual expectations of several orders, conducive to coordination by means of replication.

The agents' agreement might be an exchange of formal or tacit promises. But it need not be. Even a man whose word is his bond can remove the promissory force by explicit disavowal, if not otherwise. An exchange of declarations of present intention will be good enough, even if each explicitly retains his right to change his plans later. No one need bind himself to act against his own interest. Rather, it will be in the interest of each to do just what he has led the others to expect him to do, since that action will be best for him if the others act on their expectations.

If one does consider himself bound by a promise, he has a second, independent incentive. His payoffs are modified, since he has attached the onus of promise breaking to all but one choice. Indeed, he may modify his payoffs so much by promising that the situation is no longer a coordination problem at all. For instance, the agent's promised action might become his dominant choice: he might wish to keep his promise no matter what, coordination or no coordination.
If such a strong promise is made publicly, the others will know that they must go along with the one who has promised, for they know what he will do. Such forceful promising is a way of getting rid of coordination problems, not a way of solving them.

Explicit agreement is an especially good and common means to coordination—so much so that we are tempted to speak of coordination otherwise produced as tacit agreement. But agreement (literally understood) is not the only source of concordant expectations to help us solve our coordination problems. We do without agreement by choice if we find ourselves already satisfied with the content and strength of our mutual expectations. We do without it by necessity if we have no way to communicate, or if we can communicate only at a cost that outweighs our improved chance of coordination (say, if we are conspirators being shadowed).

Schelling has experimented with coordination problems in which the agents cannot communicate. His subjects know only that they share a common understanding of their problem—for instance, they may get instructions describing their problem and stating that everyone gets the same instructions. It turns out that sophisticated subjects in an experimental setting can often do very well—much better than chance—at solving novel coordination problems without communicating. They try for a coordination equilibrium that is somehow salient: one that stands out from the rest by its uniqueness in some conspicuous respect. It does not have to be uniquely good; indeed, it could be uniquely bad. It merely has to be unique in some way the subjects will notice, expect each other to notice, and so on. If different coordination equilibria are unique in different conspicuous ways, the subjects will need to be alike in the relative importance they attach to different respects of comparison; but often they are enough alike to solve the problem.

How can we explain coordination by salience? The subjects might all tend to pick the salient as a last resort, when they have no stronger ground for choice. Or they might expect each other to have that tendency, and act accordingly; or they might expect each other to expect each other to have that tendency and act accordingly, and
act accordingly; and so on. Or—more likely—there might be a mixture of these. Their first- and higher-order expectations of a tendency to pick the salient as a last resort would be a system of concordant expectations capable of producing coordination at the salient equilibrium.

If their expectations did produce coordination, it would not matter whether anyone really would have picked the salient as a last resort. For each would have had a good reason for his choice, so his choice would not have been a last resort.

Thus even in a novel coordination problem—which is an extreme case—the agents can sometimes obtain the concordant expectations they need without communicating. An easier, and more common, case is that of a familiar coordination problem without communication. Here the agents’ source of mutual expectations is precedent: acquaintance with past solved instances of their present coordination problem.

4. Convention

Let us start with the simplest case of coordination by precedent and generalize in various ways. In this way we shall meet the phenomenon I call convention, the subject of this book.

Suppose we have been given a coordination problem, and we have reached some fairly good coordination equilibrium. Given exactly the same problem again, perhaps each of us will repeat what he did before. If so, we will reach the same solution. If you and I met yesterday—by luck, by agreement, by salience, or however—and today we find we must meet again, we might both go back to yesterday’s meeting place, each hoping to find the other there. If we were cut off on the telephone and you happened to call back as I waited, then if we are cut off again in the same call, I will wait again.

We can explain the force of precedent just as we explained the force of salience. Indeed, precedent is merely the source of one important kind of salience: conspicuous uniqueness of an equilibrium because we reached it last time. We may tend to repeat the action
that succeeded before if we have no strong reason to do otherwise. Whether or not any of us really has this tendency, we may somewhat expect each other to have it, or expect each other to expect each other to have it, and so on—that is, we may each have first- and higher-order expectations that the others will do their parts of the old coordination equilibrium, unless they have reason to act otherwise. Each one’s expectation that the others will do their parts, strengthened perhaps by replication using his higher-order expectations, gives him some reason to do his own part. And if his original expectations of some order or other were strong enough, he will have a decisive reason to do his part. So he will do it.

I have been supposing that we are given a coordination problem, and then given the same problem again. But, of course, we could never be given exactly the same problem twice. There must be this difference at least: the second time, we can draw on our experience with the first. More generally, the two problems will differ in several independent respects. We cannot do exactly what we did before. Nothing we could do this time is exactly like what we did before—like it in every respect—because the situations are not exactly alike.

So suppose not that we are given the original problem again, but rather that we are given a new coordination problem analogous somehow to the original one. Guided by whatever analogy we notice, we tend to follow precedent by trying for a coordination equilibrium in the new problem which uniquely corresponds to the one we reached before.

There might be alternative analogies. If so, there is room for ambiguity about what would be following precedent and doing what we did before. Suppose that yesterday I called you on the telephone and I called back when we were cut off. Today you call me and we are cut off. We have a precedent in which I called back and a precedent—the same one—in which the original caller called back. But this time you are the original caller. No matter what I do this time, I do something analogous to what we did before. Our ambiguous precedent does not help us.

In fact, there are always innumerable alternative analogies.
it not that we happen uniformly to notice some analogies and ignore others—those we call “natural” or “artificial,” respectively—prece­
dents would always be completely ambiguous and worthless. Every coordination equilibrium in our new problem (every other combina­
tion, too) corresponds uniquely to what we did before under some analogy, shares some distinctive description with it alone. Fortunately, most of the analogies are artificial. We ignore them; we do not tend to let them guide our choice, nor do we expect each other to have any such tendency, nor do we expect each other to expect each other to, and so on. And fortunately we have learned that all of us will mostly notice the same analogies. That is why precedents can be unambiguous in practice, and often are. If we notice only one of the analogies between our problem and the precedent, or if one of those we notice seems far more conspicuous than the others, or even if several are conspicuous but they all happen to agree in indicating the same choice, then the other analogies do not matter. We are not in trouble unless conflicting analogies force themselves on our attention.

The more respects of similarity between the new problem and the precedent, the more likely it is that different analogies will turn out to agree, the less room there will be for ambiguity, and the easier it will be to follow precedent. A precedent in which I, the original caller, called back is ambiguous given a new problem in which you are the original caller—but not given a new problem in which I am again the original caller. That is why I began by pretending that the new problem was like the precedent in all respects.

Salience in general is uniqueness of a coordination equilibrium in a preeminently conspicuous respect. The salience due to precedent is no exception: it is uniqueness of a coordination equilibrium in virtue of its preeminently conspicuous analogy to what was done successfully before.

So far I have been supposing that the agents who set the precedent are the ones who follow it. This made sure that the agents given the second problem were acquainted with the circumstances and outcome of the first, and expected each other to be, expected each other to
expect each other to be, and so on. But it is not an infallible way and not the only way. For instance, if yesterday I told you a story about people who got separated in the subway and happened to meet again at Charles Street, and today we get separated in the same way, we might independently decide to go and wait at Charles Street. It makes no difference whether the story I told you was true, or whether you thought it was, or whether I thought it was, or even whether I claimed it was. A fictive precedent would be as effective as an actual one in suggesting a course of action for us, and therefore as good a source of concordant mutual expectations enabling us to meet. So let us just stipulate that somehow the agents in the new problem are acquainted with the precedent, expect each other to be acquainted with it, and so on.

So far I have been supposing that we have a single precedent to follow. But we might have several. We might all be acquainted with a class of previous coordination problems, naturally analogous to our present problem and to each other, in which analogous coordination equilibria were reached. This is to say that the agents’ actions conformed to some noticeable regularity. Since our present problem is suitably analogous to the precedents, we can reach a coordination equilibrium by all conforming to this same regularity. Each of us wants to conform to it if the others do; he has a conditional preference for conformity. If we do conform, the explanation has the familiar pattern: we tend to follow precedent, given no particular reason to do anything else; we expect that tendency in each other; we expect each other to expect it; and so on. We have our concordant first- and higher-order expectations, and they enable us to reach a coordination equilibrium.

It does not matter why coordination was achieved at analogous equilibria in the previous cases. Even if it had happened by luck, we could still follow the precedent set. One likely course of events would be this: the first case, or the first few, acted as precedent for the next, those for the next, and so on. Similarly, no matter how our precedents came about, by following them this time we add this case to the stock of precedents available henceforth.
Several precedents are better than one, not only because we learn by repetition but also because differences between the precedents help to resolve ambiguity. Even if our present situation bears conflicting natural analogies to any one precedent, maybe only one of these analogies will hold between the precedents; so we will pay attention only to that one. Suppose we know of many cases in which a cut-off telephone call was restored, and in every case it was the original caller who called back. In some cases I was the original caller, in some you were, in some neither of us was. Now we are cut off and I was the original caller. For you to call back would be to do something analogous—under one analogy—to what succeeded in some of the previous cases. But we can ignore that analogy, for under it the precedents disagree.

Once there are many precedents available, without substantial disagreement or ambiguity, it is no longer necessary for all of us to be acquainted with precisely the same ones. It is enough if each of us is acquainted with some agreeing precedents, each expects everyone else to be acquainted with some that agree with his, each expects everyone else to expect everyone else to be acquainted with some precedents that agree with his, etc. It is easy to see how that might happen: if one has often encountered cases in which coordination was achieved in a certain problem by conforming to a certain regularity, and rarely or never encountered cases in which it was not, he is entitled to expect his neighbors to have had much the same experience. If I have driven all around the United States and seen many people driving on the right and never one on the left, I may reasonably infer that almost everyone in the United States drives on the right, and hence that this man driving toward me also has mostly seen people driving on the right—even if he and I have not seen any of the same people driving on the right.

Our acquaintance with a precedent need not be very detailed. It is enough to know that one has learned of many cases in which coordination was achieved in a certain problem by conforming to a certain regularity. There is no need to be able to specify the time and place, the agents involved, or any other particulars; no need to
be able to recall the cases one by one. I cannot cite precedents one by one in which people drove on the right in the United States; I am not sure I can cite even one case; nonetheless, I know very well that I have often seen cars driven in the United States, and almost always they were on the right. And since I have no reason to think I encountered an abnormal sample, I infer that drivers in the United States do almost always drive on the right; so anyone I meet driving in the United States will believe this just as I do, will expect me to believe it, and so on.

Coordination by precedent, at its simplest, is this: achievement of coordination by means of shared acquaintance with the achievement of coordination in a single past case exactly like our present coordination problem. By removing inessential restrictions, we have come to this: achievement of coordination by means of shared acquaintance with a regularity governing the achievement of coordination in a class of past cases which bear some conspicuous analogy to one another and to our present coordination problem. Our acquaintance with this regularity comes from our experience with some of its instances, not necessarily the same ones for everybody.

Given a regularity in past cases, we may reasonably extrapolate it into the (near) future. For we are entitled to expect that when agents acquainted with the past regularity are confronted by an analogous new coordination problem, they will succeed in achieving coordination by following precedent and continuing to conform to the same regularity. We come to expect conforming actions not only in past cases but in future ones as well. We acquire a general belief, unrestricted as to time, that members of a certain population conform to a certain regularity in a certain kind of recurring coordination problem for the sake of coordination.

Each new action in conformity to the regularity adds to our experience of general conformity. Our experience of general conformity in the past leads us, by force of precedent, to expect a like conformity in the future. And our expectation of future conformity is a reason to go on conforming, since to conform if others do is to achieve a coordination equilibrium and to satisfy one's own preferences. And
so it goes—we're here because we're here because we're here because we're here. Once the process gets started, we have a metastable self-perpetuating system of preferences, expectations, and actions capable of persisting indefinitely. As long as uniform conformity is a coordination equilibrium, so that each wants to conform conditionally upon conformity by the others, conforming action produces expectation of conforming action and expectation of conforming action produces conforming action.

This is the phenomenon I call convention. Our first, rough, definition is:

A regularity $R$ in the behavior of members of a population $P$ when they are agents in a recurrent situation $S$ is a convention if and only if, in any instance of $S$ among members of $P$,

1. everyone conforms to $R$;
2. everyone expects everyone else to conform to $R$;
3. everyone prefers to conform to $R$ on condition that the others do, since $S$ is a coordination problem and uniform conformity to $R$ is a proper coordination equilibrium in $S$.

5. Sample Conventions

Chapter II will be devoted to improving the definition. But before we hide the concept beneath its refinements, let us see how it applies to examples. Consider some conventions to solve our sample coordination problems.

1. If you and I must meet every week, perhaps at first we will make a new appointment every time. But after we have met at the same time and place for a few weeks running, one of us will say, “See you here next week,” at the end of every meeting. Later still we will not say anything (unless our usual arrangement is going to be unsatisfactory next week). We will just both go regularly to a certain place at a certain time every week, each going there to meet the other and confident that he will show up. This regularity that has gradually developed in our behavior is a convention.
In this case the convention that sets our meeting place holds in the smallest possible population: just two people. In other cases, larger populations—perhaps with changing membership—have conventional meeting places. What makes a soda fountain, coffeehouse, or bar "in" is the existence of a convention in some social circle that it is the place to go when one wants to socialize. The man in the song—"Standing on a corner with a dollar in my hand / Looking for a woman who's looking for a man"—is standing on that corner in conformity to a convention among all the local prostitutes and their customers.

(2) In my hometown of Oberlin, Ohio, until recently all local telephone calls were cut off without warning after three minutes. Soon after the practice had begun, a convention grew up among Oberlin residents that when a call was cut off the original caller would call back while the called party waited. Residents usually conformed to this regularity in the expectation of conformity by the other party to the call. In this way calls were easily restored, to the advantage of all concerned. New residents were told about the convention or learned it through experience. It persisted for a decade or so until the cutoff was abolished.

Other regularities might have done almost as well. It could have been the called party who always called back, or the alphabetically first, or even the older. Any of these regularities could have become the convention if enough of us had started conforming to it. It would have been a bit less convenient than our actual convention; if the original caller calls back, he may still remember the number and he must at least know where to find it. But the inconveniences of another convention would not have outweighed the advantage of achieving a coordination equilibrium by calling back if and only if one's partner does not.

This example illustrates the possibility that (describing actions in any natural way) a conventional regularity may specify different actions under different conditions. In this case it specifies what we would naturally call different actions for agents involved in situation S in different roles. Except for ad hoc descriptions like "action in
conformity to such-and-such regularity,” the actions conforming to a conventional regularity do not have to share any common natural description. Therefore, when we speak of a convention to do an action \( A \) in a situation \( S \), it must be understood that \( A \) may stand for an unnaturally complex action-description.

(3) If the two rowers in Hume’s boat manage somehow to fall into a smooth rhythm and maintain it for a while, they “do it by an agreement or convention, though they have never given promises to each other.” A regularity in their behavior—their rowing in that particular rhythm—persists because they expect it to be continued and they want to match their rhythms of rowing. “This common sense of interest . . . known to both . . . produces a suitable resolution and behavior” in which “the actions of each . . . have a reference to those of the other, and are performed upon the supposition that something is to be performed upon the other part.”

This convention is peculiar. It holds in a very small population for a very short time—between two people for a few minutes—and the regularity is one we would find it very hard to describe, though we can easily catch on to it. But these oddities do not detract from its conventionality.

(4) We drive in the right lane on roads in the United States (or in the left lane on roads in Britain, Australia, Sweden before 1967, parts of Austria before a certain date, and elsewhere) because we do not want to drive in the same lane as the drivers coming toward us, and we expect them to drive on the right.

There is a complication: if we do not drive on the right, the highway patrol will catch us and we will be punished. So we have an independent incentive to drive on the right, and this second incentive is independent of how the others drive. But it makes no important difference. If I expected the others to be on the left, I would be there too, highway patrol or no highway patrol. My preference for driving on the same side as the others outweighs any incentive the highway patrol may give me to drive on the right. And so it is for almost everyone else, I am sure. The highway patrol modifies the payoffs
in favor of driving on the right; but there are still two different coordination equilibria. The punishments are superfluous if they agree with our convention, are outweighed if they go against it, are not decisive either way, and hence do not make it any less conventional to drive on the right. The same goes for other considerations favoring one coordination equilibrium over the other: the fact that our cars have left-hand drive, the fact that we are mostly right-handed, and so on.

(5) If four men who camp together find that often they waste effort by covering the same ground in search of firewood, they may get fed up and agree once and for all: let Morgan look to the north, Jones to the east, Owen to the south, Griffith to the west. From that day on, each goes his proper way without further discussion. A regularity has begun by explicit agreement. At first, perhaps, it persists because each man feels bound by his promise and takes no account of the advantages of keeping it or breaking it. But years pass. They forget that they agreed. Morgan is replaced by Thomas, who never heard of the agreement and never promised anything. Yet whenever they need firewood each still goes off in his proper direction, because he knows that is how to have the ground to himself. As the force of their original promises fades away, the regularity in their behavior becomes a convention.

(6) Wanting to attend parties dressed as the others will be dressed, we wear whatever is conventional dress for the occasion; in picking our clothes we act in conformity to a convention of our social circle. By means of a conditional conventional regularity specifying the style of clothes worn in various circumstances, we satisfy our common interest in being dressed alike.

But we must distinguish two cases. If each of us wants to dress like the majority and wants everyone else to dress like the majority too, then we achieve a coordination equilibrium when we all dress alike: our regularity is a genuine convention. Suppose, however, that many of us are nasty people who want to dress like the majority but also want to have a differently dressed minority to sneer at. We still
achieve an equilibrium when we all dress alike, but it is not a coordination equilibrium: nobody wishes he himself had dressed otherwise, but the nasty ones wish that a few other people—say, their worst enemies—had dressed otherwise. The regularity whereby we achieve this equilibrium is not a genuine convention by my definition, because the element of conflict of interest prevents it from being a means of reaching a \textit{coordination} equilibrium.

It may not be obvious that our regularities of dress should not be called conventions if there are many people who want to see them violated. But when our analysis has shown us how the presence of substantial conflict makes a disanalogy between this case and other clear cases of convention, and makes an important analogy between this case and clear cases of nonconvention like the one to be examined in Chapter III.5, I think we ought to end up agreeing with the analysis even against our first impressions. If the reader disagrees, I can only remind him that I did not undertake to analyze anyone's concept of convention but mine.

(7) If we are contented oligopolists who want to maintain a uniform but fluctuating price for our commodity, we dare not make any explicit agreement on prices; that would be a conspiracy in restraint of trade. But we can come to a tacit understanding—that is, a convention—by our ways of responding to each others' prices. We might, for instance, start to follow a price leader: one firm that takes the initiative in changing prices, with due care to set a price in the range that is satisfactory to all of us.

In this example, it becomes seriously artificial to divide our continuous activity into a sequence of separate analogous coordination problems, related only by force of precedent. (The difficulty will reappear in examples [9], [10], and [11]; it was present somewhat in [3] and [4].) We can actually set or reconsider prices at any time. How long is a coordination problem? Pretend, already idealizing, that we set our prices every morning and cannot change them later in the day. Then each business day is a coordination problem. But a day is too short. Our customers take more than a day to shop around;
they compare my price for today with yours for yesterday and someone else’s for tomorrow. We are leaving out most of the coordination: coordination of one’s action on one day with another’s action on another nearby day. If, on the other hand, we take longer stretches as the coordination problems, then—contrary to the definition—everyone has time for several different choices within a single coordination problem. We might pretend that everyone starts each week by choosing a contingency plan specifying what to do in every possible circumstance during the week (a strategy in the sense of the theory of games), and then follows his plan all week without making any further choice. Then a business week is a coordination problem in which everyone makes only his one initial choice of a contingency plan. But this treatment badly misdescribes what we do; and it still leaves out the coordination between, say, my prices for Friday and yours for next Monday. A better remedy, scheduled for Chapter II.3, goes deep. We can forget about individual coordination problems; instead of saying that uniform conformity to a regularity \( R \) constitutes a coordination equilibrium in every instance of a situation \( S \), we can say approximately the same thing in terms of conditional preferences for conformity to \( R \).

(8) If Rousseau’s stag hunters stay with the hunt every time, they do so by a convention. Each stays because he trusts the others to stay as they did before, and he will eat better by staying and taking his share of the stag when it is caught.

But, less obviously, if they always split up and catch rabbits separately, that is a convention too. If the stag hunt fails unless all take part, there is no point in joining unless all the others do. Each prefers to catch rabbits if even one of the others does, and \textit{a fortiori} if all the others do. For each to catch his rabbit is not a good coordination equilibrium. But it is a coordination equilibrium nonetheless, so long as catching a rabbit is better than going off on a one-man stag hunt that is bound to fail. So rabbit catching is, by definition, a convention.

(9) On the hypothesis that each of us wants the exclusive use of
some land and that nobody ever thinks it worth the trouble to try taking over the use of some land from another, any *de facto* division of the land is a convention. Each goes on using a certain portion and keeping off the rest in the knowledge that, since others will go on using everything else, that is the only way he can meet his needs and stay out of trouble. A better convention might provide a regular way to deal with changes in the population of land users. It might be part of the convention, for instance, that when any man dies, the oldest boy not yet using land begins to use the vacant portion.

I have not called these portions of land *property*. At its simplest—say, among anarchists—the institution of property might be nothing more than a convention specifying who shall have the exclusive use of which goods. This seems to be Hume's theory of property. For us, the institution of property is more complicated; we have built it into an elaborate system of laws and institutions. We do not say that a squatter owns the land he farms, though he enjoys the exclusive use of it by a convention, since another claimant is entitled by law to call on the police to kick the squatter off. I therefore shall not define *property* as goods reserved by convention for someone's exclusive use.

(10) A medium of exchange—say, coin of the realm—has its special status by a convention among tradesmen to take it without question in return for goods and services. Some conventional media are better than others: bulky or perishable ones are bad; ones that would retain some use if the convention collapsed are good—but the inconvenience of accepting a bad medium of exchange is less than the inconvenience of refusing it when others take it, or of taking what one can neither use nor spend. Again, as in (4), there is the complication of legal sanctions. Refusal to accept legal tender makes a debt legally unenforceable. But again, such sanctions are superfluous if they agree with convention, are outweighed if they go against it, are not decisive either way, and therefore do not make our regularity any the less conventional.

I suppose we may safely define a *medium of exchange* as any good
that is conventionally accepted in some population in return for goods and services. This definition raises an annoying question: is it right to say that we have a convention to accept our media of exchange in return for goods and services? It is false to say that our convention is that we accept our media of exchange in return for goods and services. For what follows “that” does not state any convention because it is true, by definition, of any population. On the other hand, it is true to say of our media of exchange that our convention is that we accept them in return for goods and services. My question was ambiguous. It can be read opaquely or transparently.\(^6\) It is like the question whether Hegel knew that the number of planets is greater than seven. He did not know that the number of planets is greater than seven. But he did know, of the number of planets—namely nine—that it is greater than seven.

(11) A population’s common use of some one language—Welsh, say—is a convention. The Welshmen in parts of Wales use Welsh; each uses Welsh because he expects his neighbors to, and for the sake of communication he wishes to use whatever language his neighbors use.

Does he not rather wish to use whatever language his neighbors will understand? Yes; but as a fact of human nature, he and his neighbors will best understand the language they use. So the right thing to say is that he wishes to use the language they use because that is the language they will understand. It follows that this is another case of coordination over time: he wishes to use the language they have been using most over a period in the past, a period long enough for them to have become skilled in its use.

To say that he wishes to use whichever language his neighbors use is not to say that if they switch suddenly, somehow, he would wish to switch immediately. He would not wish to, because he could not; he would have to practice their new language. Besides, he could count on them to understand Welsh for a time after they had ceased to

---

use it. But probably he would wish to switch as soon as he easily could. And if it suddenly came to pass that his neighbors had been using their new language for twenty years—while he, let us say, had been sleeping like Rip Van Winkle—he would try to conform with the utmost urgency.

I do not deny, of course, that a man may prefer one language to another—say, the language of his fathers to the language of their conquerors. But that does not matter. Different coordination equilibria do not have to be equally good—only good enough so that everyone is ready to do his part if the others do. There are few who would give up communication out of piety to the mother tongue, if it came to that.

Certainly not every feature of a language is conventional. No humanly possible language relies on ultrasonic whistles, so it is not by convention that the Welshmen do not. We do not yet know exactly which features of languages are conventional and which are common to all humanly possible languages; Noam Chomsky and his school have argued that there is less conventionality than one might have thought. But so long as even two languages are humanly possible, it must be by convention that a population chooses to use one or the other.

In saying that Welshmen use Welsh by convention, I do not say it is a convention that Welshmen use Welsh. This, or something similar but more complicated, might perhaps be true by definition of “Welsh.” Rather, I say of Welsh that it is a convention among Welshmen that they use it. The difference is the same ambiguity between opaque and transparent readings that arose in (10).

If using Welsh is to be a convention, it must be a regularity in behavior. It is not, of course, a regularity that fully determines a Welshman’s behavior. He can say a variety of things, or remain silent, and he can respond to utterances in a variety of ways, and still be conforming to the conventional regularity. But that is nothing special.

No convention determines every detail of behavior. (The meeting-place convention, for instance, does not specify whether to walk or ride to the meeting place.) This convention, like any other, restricts behavior without removing all choice. There is more choice, and more important choice, in this case than in some others; but there is no difference in kind.

A convention is a regularity in behavior. I do not want to say that the users of Welsh are conforming to their convention when and only when they are rightly said to be “using Welsh.” A man lying in Welsh is using Welsh, but he is violating its convention; a man who remains silent during a conversation may be conforming to the convention although he is not using Welsh. In due course we shall see how the convention of a language may be described; here I will say only that it is a regularity restricting one’s production of, and response to, verbal utterances and inscriptions. Linguistic competence consists in part of a disposition to conform to that restriction with ease; and in part of an expectation that one’s neighbors will be likewise disposed, with a recognition of their conformity as the reason for one’s own. No doubt a child or an idiot may conform without reason; if so, he is not party to the convention and his linguistic competence is incomplete.
II | Convention Refined

1. Common Knowledge

Agreement, salience, or precedent, we have seen, can solve a coordination problem by producing a system of concordant first- and higher-order mutual expectations. We need only imagine cases to convince ourselves that higher-order expectations would be produced. But how? What premises have we to justify us in concluding that others have certain expectations, that others expect others to have certain expectations, and so on? And how is the process cut off—as it surely is—so that it produces only expectations of the first few orders?

Take a simple case of coordination by agreement. Suppose the following state of affairs—call it A—holds: you and I have met, we have been talking together, you must leave before our business is done; so you say you will return to the same place tomorrow. Imagine the case. Clearly, I will expect you to return. You will expect me to expect you to return. I will expect you to expect me to expect you to return. Perhaps there will be one or two orders more.

What is it about A that explains the generation of these higher-order expectations? I suggest the reason is that A meets these three conditions:

(1) You and I have reason to believe that A holds.
(2) A indicates to both of us that you and I have reason to believe that A holds.
(3) A indicates to both of us that you will return.

What is indicating? Let us say that A indicates to someone x that
if and only if, if x had reason to believe that A held, x would thereby have reason to believe that __. What A indicates to x will depend, therefore, on x’s inductive standards and background information.

The three main premises (1), (2), (3), together with suitable ancillary premises regarding our rationality, inductive standards, and background information, suffice to justify my higher-order expectations. Let us see how my reasoning would work.

Consider that if A indicates something to x, and if y shares x’s inductive standards and background information, then A must indicate the same thing to y. Therefore, if A indicates to x that y has reason to believe that A holds, and if A indicates to x that __, and if x has reason to believe that y shares x’s inductive standards and background information, then A indicates to x that y has reason to believe that __ (this reason being y’s reason to believe that A holds). Suppose you and I do have reason to believe we share the same inductive standards and background information, at least nearly enough so that A will indicate the same things to both of us. Then (2) applied to (3) implies:

(4) A indicates to both of us that each of us has reason to believe that you will return.

And (2) applied in turn to (4) implies:

(5) A indicates to both of us that each of us has reason to believe that the other has reason to believe that you will return.

And so on ad infinitum, since each new conclusion begins “A indicates to both of us that . . .” Note that this is a chain of implications, not of steps in anyone’s actual reasoning. Therefore there is nothing improper about its infinite length. Figure 26 is a more detailed representation of these implications in my case; those in your case could be represented similarly.

Consider next that our definition of indication yields a principle of detachment: if A indicates to x that __ and x has reason to
I believe that you share my inductive standards and background information.

\[ A \text{ indicates to me that you have reason to believe that } A \text{ holds.} \]

\[ A \text{ indicates to me that you will return.} \]

\[ A \text{ holds.} \]

\[ A \text{ indicates to me that I have reason to believe that you will return.} \]

\[ A \text{ holds.} \]
believe that \( A \) holds, then \( x \) has reason to believe that ____ . Premise (1) applied in this way to (3) implies:

\((3')\) Each of us has reason to believe that you will return.

Premise (1) applied to (4) implies:

\((4')\) Each of us has reason to believe that the other has reason to believe that you will return.

Premise (1) applied to (5) implies:

\((5')\) Each of us has reason to believe that the other has reason to believe that the first has reason to believe that you will return.

And so on, for the whole infinite sequence we considered above. I am still not talking about anyone's actual reasoning or what anyone actually does believe. But the only actual reasoning needed now is reasoning to convert these iterations of “has reason to believe” to the corresponding iterations of “does believe.” For that we need ancillary premises about rationality.

Anyone who has reason to believe something will come to believe it, provided he has a sufficient degree of rationality. So according to (3'), if we both have a sufficient degree of rationality, then it will come to be that

\((3'')\) Each of us expects that you will return.

According to (4'), if each of us has reason to ascribe a sufficient degree of rationality to the other, then each has reason to expect that the other expects that you will return. If, in addition, we both have a sufficient degree of rationality, then it will come to be that

\((4'')\) Each of us expects that the other expects that you will return.

According to (5'), if each of us has reason to expect that the other has reason to ascribe a sufficient degree of rationality to him, then
each has reason to expect that the other has reason to expect that he expects that you will return. If, in addition, each of us has reason to ascribe a sufficient degree of rationality to the other, then each has reason to expect that the other expects that he expects that you will return. And if, in addition, we both have a sufficient degree of rationality, then it will come to be that

\( (5'') \) Each of us expects that the other expects that he expects that you will return.

And so on. Each term of the sequence \((3'), (4'), (5') \ldots \), together with sufficient rationality, reason to ascribe sufficient rationality, etc., guarantees formation of the corresponding first- or higher-order expectation. But the degrees of rationality we are required to have, to have reason to ascribe, etc., obviously increase quickly. That is why expectations of only the first few orders are actually formed. The generating process stops when the ancillary premises give out.

This completes our example of a state of affairs which produces higher-order expectations. I take this example to be typical; all the higher-order expectations involved in sustaining conventions, and more or less all we ever have, seem to be produced in this way.

Let us say that it is *common knowledge* in a population \( P \) that ___ if and only if some state of affairs \( A \) holds such that:

1. Everyone in \( P \) has reason to believe that \( A \) holds.
2. \( A \) indicates to everyone in \( P \) that everyone in \( P \) has reason to believe that \( A \) holds.
3. \( A \) indicates to everyone in \( P \) that ____.

We can call any such state of affairs \( A \) a *basis* for common knowledge in \( P \) that _____. \( A \) provides the members of \( P \) with part of what they need to form expectations of arbitrarily high order, regarding sequences of members of \( P \), that _____. The part it gives them is the part peculiar to the content _____. The rest of what they need is what they need to form *any* higher-order expectations in the way we are considering: mutual ascription of some common inductive standards
and background information, rationality, mutual ascription of rationality, and so on.

Let us return to our example and consider the state of affairs $A$ more completely. Suppose that as part of $A$ we manifest our conditional preferences for returning to the meeting place. Then $A$ may also indicate to us that we both have such preferences. If so, $A$ can serve as a basis not only for common knowledge that you will return, but also as a basis for common knowledge that each of us prefers to return if the other does. Suppose also that as part of $A$ we somehow manifest a modicum of rationality. Then $A$ may indicate to us, and be a basis for common knowledge of, our possession of this modicum of rationality. By now $A$—our incident of agreeing to return—is generating all the higher-order expectations that contribute to our success in solving our coordination problem by means of replication.

A basis for common knowledge generates higher-order expectations with the aid of pre-existing higher-order expectations of rationality. Can these themselves be generated by some basis for common knowledge? Yes, because all the higher-order expectations of rationality needed to generate an $n$th-order expectation are themselves of less than $n$th-order. What cuts off the generation of higher-order expectations is the limited amount of rationality indicated by any basis—not any difficulty in generating higher-order expectations of as much rationality as is indicated by a basis.

Agreement to do one's part of a coordination equilibrium is a basis for common knowledge that everyone will do his part. Salience is another basis for common knowledge that everyone will do his part of a coordination equilibrium; but it is a weaker basis, in general, and generates weaker higher-order expectations, since the salience of an equilibrium is not a very strong indication that agents will tend to choose it. Precedents also are a basis for common knowledge that everyone will do his part of a coordination equilibrium; and, in particular, past conformity to a convention is a basis for common knowledge of a tendency to go on conforming. Consider a conventional regularity $R$ in a population $P$. Everyone in $P$ has reason to
believe that members of $P$ have conformed to $R$ in the past. The fact that members of $P$ have conformed to $R$ in the past indicates to everyone in $P$ that everyone in $P$ has reason to believe that members of $P$ have conformed to $R$ in the past. And the fact that members of $P$ have conformed to $R$ in the past indicates to everyone in $P$ that they will tend to do so in the future as well.

For example, drivers in the United States have hitherto driven on the right. All of us have reason to believe that this is so. And the fact that this is so indicates to all of us that all of us have reason to believe that drivers in the United States have hitherto driven on the right and also that drivers in the United States will tend to drive on the right henceforth.

Our defining conditions for the existence of a convention consist of a regularity in behavior, a system of mutual expectations, and a system of preferences. I propose to amend the definition: not only must these conditions be satisfied, but also it must be common knowledge in the population that they are. Our amended definition is:

A regularity $R$ in the behavior of members of a population $P$ when they are agents in a recurrent situation $S$ is a convention if and only if it is true that, and it is common knowledge in $P$ that, in any instance of $S$ among members of $P$,

(1) everyone conforms to $R$;
(2) everyone expects everyone else to conform to $R$;
(3) everyone prefers to conform to $R$ on condition that the others do, since $S$ is a coordination problem and uniform conformity to $R$ is a coordination equilibrium in $S$.

Thus there is to be some state of affairs $A$ (such that $A$ holds, everyone in $P$ has reason to believe that $A$ holds, and $A$ indicates to everyone in $P$ that everyone in $P$ has reason to believe that $A$ holds) which indicates to everyone in $P$ that members of $P$ conform to $R$, that they expect each other to conform to $R$, and that they have prefer-
ences which make uniform conformity to $R$ a coordination equilibrium.

One reason to amend the definition of convention is simply that we want to write into the definition all of the important features common to our examples, and common knowledge of the relevant facts seems to be one such feature. There is another reason: the amendment helps to deal with certain odd cases, regularities which seem intuitively unlike clear cases of convention but which would have qualified as conventions under the unamended definition.

Suppose everyone drives on the right because he expects everyone else to drive on the right and he wants to prevent collisions. But suppose no one gives anyone else credit for intelligence equal to his own. Everyone holds this false belief (call it $f$): “Except for myself, everyone drives on the right by habit, for no reason, and would go on driving on the right no matter what he expected others to do.” This is a case of convention under the unamended definition, despite the false beliefs; but I think it ought to be excluded. It cannot be a case of convention under the amended definition (unless we are extremely irrational); for if it is, there is some state of affairs which we have reason to believe holds and which indicates to us that $f$ is false. This case is only the first of a sequence. Suppose next that no one really has the false belief $f$, but everyone falsely ascribes it to everyone else. This too cannot be a case of convention under the amended definition (unless we are extremely irrational); for if it is, there is some state of affairs which we have reason to believe holds and which indicates to us that everyone has reason to disbelieve $f$. And so on. The cases become more and more unlikely, but no less deserving of exclusion; the amended definition continues to exclude them (given a sufficiently strong assumption of rationality—stronger and stronger assumptions of rationality are needed as we go on).

By now one might guess that common knowledge is the only possible source of higher-order expectations. But it is not; there is a general method for producing expectations of arbitrarily high order in isolation. For instance, I can acquire an isolated fourth-order
expectation as follows. Suppose I am a resident of Ableton and I believe everything printed in the *Ableton Argus*. Today’s *Argus* prints this story:

The *Bakerville Bugle* is totally unreliable; what it prints is as likely to be false as to be true. Yet the residents of *Bakerville* believe everything in it. Today’s *Bugle* printed this story:

The *Charlie City Crier* is totally unreliable; what it prints is as likely to be false as to be true. Yet the residents of *Charlie City* believe everything in it. Today’s *Crier* printed this story:

The *Dogpatch Daily* is totally unreliable; what it prints is as likely to be false as to be true. Yet the residents of Dogpatch believe everything in it. Today’s *Daily* printed this story:

Tomorrow it will rain cats and dogs.

I should not expect it to rain cats and dogs. I should not expect the residents of Dogpatch to expect it to rain cats and dogs. I should not expect the residents of Charlie City to expect the residents of Dogpatch to expect it to rain cats and dogs. But I should expect the residents of Bakerville to expect the residents of Charlie City to expect the residents of Dogpatch to expect it to rain cats and dogs. In other words, I should have a fourth-order expectation that it will rain cats and dogs, without any corresponding lower-order expectations that it will. Obviously, the method would have worked for an arbitrarily long sequence of newspapers; the sequence could have repeated, provided no two adjacent terms were the same. I do not claim that this method of generating isolated higher-order expectations is of much practical importance; it merely establishes the possibility.

2. Knowledge of Conventions

Suppose it is common knowledge in a population $P$ that some state of affairs $B$ holds. Then everyone in $P$ has reason to expect it to be
common knowledge in $P$ that $B$ holds. For by definition of common knowledge, there is some state of affairs $A$ such that:

1. Everyone in $P$ has reason to believe that $A$ holds;
2. $A$ indicates in $P$ that everyone in $P$ has reason to believe that $A$ holds;
3. $A$ indicates in $P$ that $B$ holds.

From (1) and (2) we may infer:

4. Everyone in $P$ has reason to believe that everyone in $P$ has reason to believe that $A$ holds.

From (2) by itself we may infer:

5. Everyone in $P$ has reason to believe that $A$ indicates in $P$ that everyone in $P$ has reason to believe that $A$ holds.

Likewise from (3) we may infer:

6. Everyone in $P$ has reason to believe that $A$ indicates in $P$ that $B$ holds.

And from (4), (5), and (6) we may infer that everyone in $P$ has reason to believe that there is a state of affairs $A$ which satisfies conditions (1), (2), and (3).

So if a convention, in particular, holds as an item of common knowledge, then to belong to the population in which that convention holds—to be party to it—is to know, in some sense, that it holds. If a regularity $R$ is a convention in population $P$, then it must be true, and common knowledge in $P$, that $R$ satisfies the defining conditions for a convention. If it is common knowledge that $R$ satisfies them, then everyone in $P$ has reason to believe that it is true, and common knowledge in $P$, that $R$ satisfies them; which is to say that everyone in $P$ must have reason to believe that $R$ is a convention.

This is not to say that a party to the convention has any special, infallible way of acquiring his knowledge. But he must have acquired it somehow, in an ordinary way, in order to be one of those among
whom the convention holds. Discovery of the convention is the principal part of one's initiation into it.

Consider the conventions of language, whatever they may be. Anyone who is a member of a population $P$, and party to its conventions of language, must know what those conventions are. If any regularity $R$ is in fact a convention of language in $P$, any normal\(^1\) member of $P$ must have reason to believe that $R$ satisfies the defining conditions for a convention.

Here is a vindication of sorts for Stanley Cavell’s doctrine that a native speaker has no need of evidence to justify that he says about what he would say. Take a philosopher who claims we would not call an action voluntary if it were not abnormal. He need not cite occasions on which people have failed to call normal actions voluntary, for he is a native speaker of the language he is telling us about. Why is he excused? Not because he, as a native speaker, has some peculiar and infallible way of acquiring his knowledge of his language. And not because his knowledge of what we would say is not real knowledge, as Cavell seems to think when he says, “the native speaker can rely on his own nose; if not, there would be nothing to count.” For the man who says what we would say is not just speaking for himself.\(^2\)

Rather, it is because the knowledge Cavell has in mind is the speaker’s knowledge of conventions to which he himself is a party. When Cavell speaks of our knowledge of “what we would say,” I take it he means our knowledge of what we could say—could say without violating our conventions of language. He does not mean our knowledge of what we would say in order to provide our audience with the information they want; of what we would say in order not to be rude or boring; of what we would say in order not to divulge trade secrets; of what we would say in order not to twist our tongues.

\(^1\)Not counting children and the feeble-minded, who may conform to $R$ without expecting conformity and without preferring to conform conditionally upon the conformity of others.

Once we have acknowledged that someone is a native speaker of our language, we have already granted that he is party to our conventions. Therefore he knows what those conventions prescribe; he knows “what we would say” in the sense in question. If we turn around and ask him to produce evidence for what he says about what we would say, we challenge his status as a native speaker and as a party to the conventions. We do not challenge some further status he might claim as an authority on the conventions as well as a party to them. He has evidence—perfectly ordinary evidence. But if we ask him to show it, we question his membership in the linguistic community to which he purports to belong. It makes no sense both to demand evidence for what he says about conventions and to take for granted that he is party to those conventions.

This vindication of Cavell’s doctrine is a poor sort of vindication, however, because our knowledge of our conventions—that minimum of knowledge everyone has in virtue of his own participation—may be quite a poor sort of knowledge:

1) It may be merely potential knowledge. We must have evidence from which we could reach the conclusion that any of our conventions meets the defining conditions for a convention, but we may not have done the reasoning to reach the conclusion. If asked whether something is a convention, we might give a snap judgment instead of evaluating our evidence; so we might get the wrong answer.

2) It may be irretrievably nonverbal knowledge. We recall the rowers in Hume’s boat, example (3) of Chapter I.5. If I am one of the rowers who row in a certain rhythm by a tacit and temporary convention, I have evidence that we have a convention to row in that rhythm. Our success in rowing in that rhythm for the last few strokes is evidence by which I arrive at my expectation that you will continue to row thus; that you prefer to row thus if I do; and that you expect me to go on rowing thus. And it is evidence that you observe this same evidence. I can use such evidence, I can expect you to use it, and so on; but I cannot describe it. I cannot say how we are rowing—say, one stroke every 2.3 seconds—but I can keep on rowing that way; I can tell whether you keep on rowing that way; later, I could
probably demonstrate to somebody what rhythm it was; I would be surprised if you began to row differently; and so on. Now there is a description that can identify the way we are rowing. We take 1.4 ± .05 seconds for the stroke and .9 ± .1 for the return, exerting a peak force of 70 ± 10 pounds near the beginning of each stroke, moving the oars from 32° ± 6° forward to 29° ± 4° back, and so on, in as much detail as you please. But, as we row, we have no use for this sort of description. We can neither give it nor tell whether it is true if somehow it is given. We would need instruments, and even if we had them we could not go on rowing as we were while we took the measurements.

Like it or not, we have plenty of knowledge we cannot put into words. And plenty of our knowledge, in words or not, is based on evidence we cannot hope to report. Our beliefs are formed under the influence of impressions left by a body of past experience, but it is only occasionally that these impressions allow us to report the experience that created them. You probably believe that Kamchatka exists. Your belief is justified, for it is based on evidence: mostly your exposure to various books and to incidents that confirm the reliability of such books. Try, then, to make a convincing case for the existence of Kamchatka by reporting parts of your experience. There is no reason why our knowledge of our conventions should be especially privileged. Like any other knowledge we have, it can be tacit, or based on tacitly known evidence, or both.

(3) It may be knowledge confined to particular instances, taken one at a time. A regularity is conventional in virtue of certain general expectations and preferences regarding conformity to it. But these will not have to be general in sensu composito; generality in sensu diviso will suffice.

The distinction, Abelard's, is this. If I expect every driver to keep right, in sensu composito, then I have one expectation with general content: I expect that every driver will keep right. It does not follow that if Jones is a driver, I expect that he will keep right, for I might not realize he is a driver. Indeed, I might even realize that Jones is a driver and still not expect that he will keep right, for I might
fail to draw the proper conclusion from my general expectation. If, on the other hand, I expect every driver to keep right, in sensu _diviso_, then I have many expectations, each with _nongeneral_ content. I expect _of_ Jones, a driver, that _he_ will keep right. Of Morgan, too. And so on, for all the drivers there are. I need not know that Jones, Morgan, and the rest are all the drivers there are; I might falsely believe there are other drivers who do not keep right. Or I might altogether lack the general concept of a driver. Generality _in sensu composito_ and generality _in sensu diviso_ are compatible and often coexist; but it is possible to have either one without the other.

Generality _in sensu diviso_ is problematic because expectation and the like apply fundamentally to states of affairs. If I expect that each driver will keep right, I do expect a state of affairs: each driver will keep right. But if I expect, _of_ each driver, that _he_ will keep right, what states of affairs do I expect? _"He will keep right"_ does not specify any state of affairs until the pronoun has been replaced by some sort of description—verbal, pictorial, or otherwise—of the person in question. Suppose the description, “the driver of the puce Cadillac ahead of me,” fits _x_. Then I can expect _of_ _x_ that _he_ will keep right by having an expectation which attaches to _x_ through that description of him: I expect that the driver of the puce Cadillac ahead of me will keep right. In that case, there is a state of affairs I expect. But not just any description of _x_ will do. Suppose, unknown to me, _x_ happens to be the chief of police and also the town drunk. I do not expect _of_ _x_ that _he_ will keep right just because I expect that the chief of police will keep right. I do not fail to expect _of_ _x_ that _he_ will keep right just because I do not expect that the town drunk will keep right. My expectation needs to be attached to _x_ by a description of some special sort; and it is hard to say which descriptions will do, and why.\(^3\)

Consider the general case: I expect every member of _P_ involved

---

with me in an instance of $S$ to conform to $R$. We have two universal quantifications: one over instances of $S$, another over members of $P$ involved in any one instance. It is possible, of course, for my expectation to be general \textit{in sensu diviso} over instances of $S$, but general \textit{in sensu composito} over agents in any one instance. That is, it might be that I expect of any instance of $S$ in which I am involved that everyone in it will conform to $R$. (Of course it is not possible for my expectation to be general \textit{in sensu composito} over instances and \textit{in sensu diviso} over agents.)

The same distinction between kinds of generality applies to other attitudes. Take our conditional preferences for conformity to convention—say, my preference for conforming to $R$ in instances of $S$ among members of $P$. If my preference is general \textit{in sensu composito} over instances of $S$, then I prefer the state of affairs in which I conform to $R$ whenever I am involved in an instance of $S$ (among members of $P$ who conform to $R$) to the state of affairs in which I sometimes fail to conform to $R$ when I am involved in an instance of $S$ (among members of $P$ who conform to $R$). But if my preference is general \textit{in sensu diviso}, then for any instance of $S$ (among members of $P$) in which I am involved, I prefer the state of affairs in which I and the others conform to $R$ in that instance to the state of affairs in which the others conform to $R$ in that instance but I do not. Again the two kinds of generality can and often do coexist, but they are independent.

Which kind of generality over instances of $S$ is wanted in the definition of convention? I should say: whichever kind it is that ensures the agent's ability to apply his general attitudes to the instance at hand. And that is a limited generality \textit{in sensu diviso}. Whenever the agent finds himself in an instance of $S$ among members of $P$, he must expect the others to conform to $R$ \textit{in that instance}, prefer to conform to $R$ if they do \textit{in that instance}, and so on, in order that he may have reason to conform to $R$ himself. Attitudes general \textit{in sensu composito} would be a likely and welcome addition, and could serve as a source of attitudes general \textit{in sensu diviso}. But they would
not be enough by themselves; the agent would have to be able to recognize instances of $S$ and derive the proper particular attitudes. If he did, his attitudes—or at least his propensity to acquire attitudes—would be general in sensu diviso.

We can imagine how a convention $R$ regarding action in $S$ might hold in a population $P$ of creatures incapable of having any attitudes general in sensu composito. They learn from experience not by coming to believe generalizations, but by acquiring propensities to come up with the right particular beliefs regarding any new case that is presented in sufficient detail. They are exposed to precedents: what we would call (but they could not) instances of $S$, in which outcomes satisfactory to all concerned were reached by what we would call (but they could not) conformity to $R$. Thereafter, whenever one of them is presented with a new instance of $S$, even one not quite like any precedent, he has all the proper attitudes regarding that instance.

He expects each other agent involved to do something we would call (but he could not) conformity to $R$. Considering any two outcomes, he has a preference; and his system of preferences between outcomes is such that there is a coordination equilibrium in which all concerned conform to $R$. Finally, it is common knowledge among members of $P$ that these attitudes are present in each who is involved in this particular instance of $S$. There is a state of affairs $A$ such that, for this or any other instance of $S$, for each one involved therein, $A$ indicates that he has the appropriate attitudes in that instance.

Such a creature has a convention and knows it to this extent: given any instance of $S$, he knows how he and each of his fellows would act therein (namely, in some way that we would call conformity to $R$). And he knows that they do so by convention; that is, given any of the defining conditions of convention as applied to a given agent in the given situation, he knows the condition is satisfied. But he cannot think of more than one instance—the given one—at a time. He has no general concept of an instance of $S$, of a member of $P$, or of an action in conformity to $R$.

Suppose we who do generalize want to exploit this creature's
knowledge of his convention, in order to give a general description of that convention. We will have to proceed by trial and error, thinking up hypotheses and trying them out on him in one (well-chosen) instance after another until we think we can predict his response to any future instance.

Even we who could know our own conventions generally in sensu composito might happen to know them only generally in sensu diviso. If we wanted to know them generally in sensu composito as well, we would have to resort to the same sort of trial and error, with ourselves as subjects. Our data about instances of our own conventions would be reliable. But our general hypotheses to systematize those data would be ordinary tentative hypotheses with no privileged status.

3. Alternatives to Conventions

One of my defining conditions for the conventionality of a regularity \( R \) regarding choice of action by agents in a situation \( S \) has been:

\[
\text{In any instance of } S \text{ among members of } P, \text{ everyone prefers to conform to } R \text{ on condition that the others do, since } S \text{ is a coordination problem and uniform conformity to } R \text{ is a coordination equilibrium in } S.
\]

In the discussion of example (7) in Chapter I.5, we found this condition unsatisfactory whenever we had coordination between actions in nearby instances of \( S \) within some continuous activity, not just coordination between actions in any one instance of \( S \). And we saw that the remedy was not to take longer stretches of activity as our coordination problems, for longer stretches are not coordination problems. Here I shall state new conditions that differ from the old one only by not requiring our activity to be chopped up into self-contained coordination problems. Our new conditions will not imply that \( S \) is a self-contained problem of interdependent decision, in
which each agent involved makes one choice of action and the outcome for each depends on the actions of all; but it will imply that if $S$ is that, then $S$ is a coordination problem and uniform conformity to $R$ is a coordination equilibrium in $S$. The special case of a sequence of coordination problems will be covered as before; but we shall find that we have taken care of the other cases at the same time.

First, we require that each agent involved in an instance of $S$ prefers to conform to $R$ conditionally upon conformity by the others involved with him in $S$. He prefers uniform conformity to $R$ to any combination of actions in which the rest conform and he does not. If $S$ is a self-contained problem of interdependent decision, this first requirement makes uniform conformity to $R$ an equilibrium. Otherwise, uniform conformity is not an equilibrium but something closely resembling one.

Second, we require that all agents involved have approximately the same preferences regarding combinations of their actions, so that $S$ is a situation in which coincidence of interests predominates. In particular, we require that all share the conditional preference of each for his conformity to $R$. That is, just as I prefer to conform if you and the others do, you also prefer me to conform if you and the others do. Taking this and the first condition together: each prefers that everyone conform to $R$, on condition that at least all but one conform to $R$, whether that one is himself or someone else. If $S$ is a self-contained problem of interdependent decision, this second requirement makes uniform conformity to $R$ a proper coordination equilibrium.

Finally, we require that there is a second possible regularity $R'$ (regarding choice of action by agents in $S$) which meets the same conditions we are imposing on $R$. We call $R'$ an alternative to $R$. It is enough to require $R'$ to meet the first and second conditions imposed on $R$. The third is automatic: if $R$ has $R'$ as an alternative, then $R'$ has $R$ itself as an alternative. If $S$ is a self-contained problem of interdependent decision, this last requirement makes uniform conformity to $R'$ a second proper coordination equilibrium. Thereby
it ensures that $S$ meets the last condition defining a coordination problem: possession of two or more proper coordination equilibria.

Recall the discussion in Chapter I.2 of the triviality of any situation with a unique coordination equilibrium and predominantly coincident interests. We are now in a better position to describe this triviality: common knowledge of rationality is all it takes for an agent to have reason to do his part of the one coordination equilibrium. He has no need to appeal to precedents or any other source of further mutual expectations.

So far we have protected convention against this triviality by requiring $S$ to be a coordination problem and hence, by definition, to have more than one proper coordination equilibrium. Now that we no longer require $S$ to be a coordination problem, our requirement for an alternative continues the same policy. In fact, whenever $S$ is a self-contained problem of interdependent decision, we have made no change at all.

Our new condition does serve to make evident one property of conventions that was not emphasized before: there is no such thing as the only possible convention. If $R$ is our actual convention, $R$ must have the alternative $R'$, and $R'$ must be such that it could have been our convention instead of $R$, if only people had started off conforming to $R'$ and expecting each other to. This is why it is redundant to speak of an arbitrary convention. Any convention is arbitrary because there is an alternative regularity that could have been our convention instead. A convention that is not arbitrary, so to speak, is a regularity whereby we achieve unique coordination equilibria. Because it is not arbitrary, it does not have to be conventional either. We would conform to it simply because that is the best thing to do. No matter what we had been doing in the past, a failure to conform to the "nonarbitrary convention" could only be a strategic error (or compensation for someone else's anticipated strategic error, or compensation for someone else's anticipated compensation, etc.).

When we try to state the requirement for an alternative more carefully, a question arises. $R$ and $R'$ are supposed to be different,
which is to say that action in conformity to $R$ (by an agent in $S$) is not also in conformity to $R'$, and vice versa. But different always, or different sometimes? After all, instances of $S$ do not have to be exactly alike. They merely have to be analogous, to fall under some common description that is natural enough to allow common knowledge of a propensity to extrapolate from some instances of $S$ to others. So action in conformity to $R$ might also be in conformity to $R'$ for some agents in some instances of $S$, though not for all.

It is not good enough to require an alternative $R'$ differing from $R$ merely to the extent of being incompatible with $R$ for some, or even for all, agents in some possible instances. Suppose $S$ occurs in a frequent version, shown in Figure 27, and in a rare version, shown in Figure 28. (We neglect any further differences between instances of a version.) $S$ is trivial in one version but not in the other. Let $R$ be the regularity of doing $R1$ or $C1$ (in the frequent version) or $R3$ or $C3$ (in the rare one). I take it that $R$ ought not to qualify as a convention, since it is trivial in most instances of $S$. But it would qualify as a convention if we counted $R'$ as its alternative, where $R'$ is the regularity of doing $R1$ or $C1$ or $R4$ or $C4$. $R'$ is an eligible regularity that is incompatible with $R$ in some instances of $S$.

It would be better to require an alternative $R'$ that is uniformly incompatible with $R$, incompatible for every agent in every instance of $S$. Now $R$ in the example above is disqualified. Its only uniformly incompatible alternative would be $R''$, the regularity of doing $R2$ or
C2 or R4 or C4. But $R''$ is not an alternative to $R$, since $R''$ usually fails to meet the requirement of conditional preference for conformity. In instances of the frequent version of $S$, no one wants to conform to $R''$ even if his partner does.

If every instance of $S$ is a coordination problem, and if uniform conformity to $R$ is always a proper coordination equilibrium, then we can find another proper coordination equilibrium in every instance of $S$. Hence the regularity $R'$ whereby one does his part of a selected second proper coordination equilibrium in every instance of $S$ is an alternative to $R$, and $R$ and $R'$ are uniformly incompatible.

If we prefer, however, we do not have to require a uniformly incompatible alternative to $R$. As a (seemingly) weaker version, we could just require that for every instance of $S$, there is a suitable regularity $R'$ which is incompatible with $R$ (for everyone involved) in that instance. Partially incompatible alternatives to $R$ are good enough if there are enough of them. The two versions are not really different. The strong version implies the weak version directly; and the weak version implies the strong version indirectly, since we can always get a uniformly incompatible alternative by patching together pieces of partially incompatible ones.

Therefore we might replace our original condition by two new ones. This one:

In any instance of $S$ among members of $P$, everyone has approximately the same preferences regarding all possible combinations of actions.

together with either this one (strong version):

There is some possible regularity $R'$ in the behavior of members of $P$ in $S$, such that no one in any instance of $S$ among members of $P$ could conform both to $R'$ and to $R$, and such that in any instance of $S$ among members of $P$, everyone would prefer that everyone conform to $R'$, on condition that at least all but one conform to $R'$.

or this one (weak version):

4
In any instance of $S$ among members of $P$, there is some possible regularity $R'$ in the behavior of members of $P$ in $S$, such that no one in that instance of $S$ could conform both to $R'$ and to $R$, and such that everyone would prefer that everyone conform to $R'$, on condition that at least all but one conform to $R'$.

I see nothing to choose between the two versions, and I choose the strong version for no good reason.

When $S$ is a self-contained problem of interdependent decision, our new conditions agree with the original condition. But they do not require $S$ to be self-contained. If not—as in my example of price setting, with $S$ taken as a stretch of business activity long enough to include several pricing decisions—the new conditions are a natural extension of the original condition.

Let $R$ be a convention regarding behavior in a coordination problem $S$; and let $R'$ be another possible regularity, partially or uniformly incompatible with $R$, which would solve $S$. But suppose the coordination equilibrium we would reach by conforming to $R'$ is much worse than the one we reach by conforming to $R$—so much worse, in fact, that it is only slightly preferred to some of the outcomes that are not coordination equilibria. Then do we really want to call $R'$ a possible alternative convention? And do we want to say that $R'$ contributes to the arbitrariness and conventionality of $R$? Perhaps not. Fortunately, our definition as it stands is likely to disqualify this $R'$ as an alternative to $R$. It may be true that:

In any instance of $S$ among members of $P$, everyone would prefer that everyone conform to $R'$, on condition that at least all but one conform to $R'$.

But it may not be true as an item of common knowledge. Our weak conditional preferences for conformity to $R'$ may well fail to be indicated by any state of affairs $A$ which we all believe to hold and which indicates to us that it holds. But of course we still require the satisfaction of the conditions to be common knowledge in $P$. And rightly: if our conditional preference for conformity to $R'$ existed,
but not as an item of common knowledge, \( R' \) could not have sustained itself in the way a convention does, so it is not true that \( R' \) could have been our convention instead of \( R \).

Consider a convention establishing some meeting place. Its alternatives would be the possible regularities whereby we would meet at other places. Some places are better than others. Some are so bad we would forgo meeting rather than go there. Considering worse and worse places, we come to the point where conditional preference for conformity fails: some of us would not want to go to the place even if the others were there. Common knowledge of conditional preference fails sooner: there are places good enough that each would want to go there if the others were there, but not good enough that we could count on each other to want to go there if the others did, count on each other to count on each other to want to go there if the others did, and so on. These places do not provide alternatives to our convention, so they do not contribute to the conventionality of our meeting place.

Or consider the conventions of our language. Their alternatives are the conventions of other possible languages. But how about a hypothetical language—or shall we call it a cipher?—so clumsy that even after any amount of practice we would still take minutes of paper-and-pencil calculation to construct or construe its easiest sentences? All of us might find even that language better than none, worth learning to use among others who used it. But if it were not common knowledge that we would, this language would not be among the alternatives that make our actual language conventional.

If the only alternatives to \( R \) were of this deficient sort, \( R \) would not be a convention. Neither would it have been a convention under the original condition requiring that \( S \) be a coordination problem, since that would not have been common knowledge either. Nor should it be called a convention. If the alternatives to \( R \) are such an inconspicuous feature of the situation, \( R \) seems almost as trivial as if they were not there at all.

Can \( R' \) be an alternative to \( R \) if the idea of acting in conformity
to \( R' \) has never occurred to anybody in \( P \)? The principle is the same: the unfamiliarity disqualifies \( R' \) if and only if it interferes with common knowledge of conditional preference for conformity to \( R' \). It may or may not. Because the expectations and preferences mentioned in the definition of convention need only be general in \( \textit{sensu diviso} \), it does not matter if we have no general concept of action in conformity to \( R' \). In fact, I argued in the last section that it would be all right if we had no general concept even of action in conformity to our actual convention. The unfamiliarity would matter, however, if it led one to fail to appreciate the advantage of some action in conformity to \( R' \) when presented with a particular instance of \( S \) in which the others did conform to \( R' \); or if it led to a failure of common knowledge that one \textit{would} appreciate that advantage.

Again there is no disagreement with the original condition, where it applies. If the only alternatives to \( R \) are disqualified by their unfamiliarity, it would not be common knowledge that \( S \) was a coordination problem. So \( R \) would not be a convention under either condition, though it might become one whenever the members of \( P \) became acquainted with the possibility of acting in conformity to \( R' \).

What is not conventional among narrow-minded and inflexible people, who would not know what to do if others began to behave differently, may be conventional among more adaptable people. What is not conventional may become conventional when news arrives of aliens who behave differently; or when somebody invents a new way of behaving, even a new way no one adopts. When children and the feeble-minded conform to our conventions, they may not take part in them as conventions, for they may lack any conditional preference for conformity to an alternative; or they may have the proper preferences, but not as an item of common knowledge. I find these corollaries of our analysis of convention neither welcome nor unwelcome. The analysis is settling questions hitherto left open.

If it seems reasonable to exclude alternatives that are too unsatisfactory or unfamiliar, as not contributing to the arbitrariness of a
convention, we have a new reason to require common knowledge. For it is by means of our common-knowledge requirement that we can exclude them without doing so ad hoc.

4. Degrees of Convention

We have confined our attention to perfect cases of convention, to which our definition applies without exceptions. But we cannot hope to find many perfect specimens in reality. It is time to be less strict, to allow for conventions that meet the present definition only for the most part or with high probability. Let us assemble the definition as amended so far:

A regularity $R$ in the behavior of members of a population $P$ when they are agents in a recurrent situation $S$ is a convention if and only if it is true that, and it is common knowledge in $P$ that, in any instance of $S$ among members of $P$,

1. everyone conforms to $R$;
2. everyone expects everyone else to conform to $R$;
3. everyone has approximately the same preferences regarding all possible combinations of actions;
4. everyone prefers that everyone conform to $R$, on condition that at least all but one conform to $R$;
5. everyone would prefer that everyone conform to $R'$, on condition that at least all but one conform to $R'$,

where $R'$ is some possible regularity in the behavior of members of $P$ in $S$, such that no one in any instance of $S$ among members of $P$ could conform both to $R'$ and to $R$.

We can count many explicit and implicit universal quantifications; we want to find a reasonable way of relaxing some or all of these to almost-universal quantifications.

The common-knowledge requirement involves universal quantifications over $P$ (see the definition of common knowledge). We need
not allow any exception to these; anyone who might be called an exception might better be excluded from $P$. It follows, however, that most of our specifications of a population in which a convention holds will be only approximately correct.

There is no harm in allowing a few abnormal instances of $S$ which violate some or all of clauses (1)–(5). So we replace “in any instance of $S$ among members of $P$” by “in almost any instance of $S$ among members of $P$.” If we ever want more precision, we can replace it by “in a fraction of at least $d_0$ of all instances of $S$ among member of $P$” with $d_0$ set slightly below one.

Nor is there any harm in allowing some, or even most, normal instances of $S$ to contain a few abnormal agents who may be exceptions to the initial universal quantifications in some or all of clauses (1)–(5). So we replace each initial “everyone” by “almost everyone” or by “everyone in a fraction of at least $d_i$ of all those involved,” with each $d_i$ set slightly below one. (We have $d_1$ for clause (1), $d_2$ for clause (2), $d_3$ for clause (3), and $d_4$ for clauses (4) and (5)—the same for both, since they are intended to be parallel.)

If we allow there to be a few agents who will not conform, we should allow the rest of the agents to know it; so “everyone else” in clause (2) should be replaced by “almost everyone else” or by “everyone else in a fraction of at least $d_1$ of all those involved.” And if we allow the agents not to expect perfect conformity, we must not make their preferences for conformity conditional upon otherwise perfect conformity; otherwise we would not guarantee that they did prefer conformity in most cases. Their preferences should be such that if enough conform, then the more the better. (So one thing we do not tolerate is a convention to which most people want there to be exceptions, however few the exceptions they want.) Clause (4) should therefore be amended again to read “prefers that any one more conform to $R$, on condition that almost everyone conform to $R$” or “prefers that any one more conform to $R$, on condition that a fraction of at least $d_4$ of all those involved conform to $R$.” Although this amendment makes clause (4) more strict rather than less, it is
unavoidable given our relaxation of clauses (1) and (2). Clause (5) should be amended in the same way to keep it parallel to (4).

We may also tolerate a few exceptions to the required incompatibility between \( R \) and its alternative \( R' \)—exceptions for most agents in a few instances of \( S \), for a few agents in most instances of \( S \), or both. We replace the incompatibility clause by "such that almost no one in almost any instance of \( S \) among members of \( P \) could conform both to \( R' \) and to \( R \)," or by "such that for a fraction of at least \( d_s \) of all pairs of an instance of \( S \) among members of \( P \) and an agent therein, the agent could not conform both to \( R' \) and to \( R \)," with \( d_s \) set slightly below one.

Our final definition is therefore:

A regularity \( R \) in the behavior of members of a population \( P \) when they are agents in a recurrent situation \( S \) is a convention if and only if it is true that, and it is common knowledge in \( P \) that, in almost any instance of \( S \) among members of \( P \),

1. almost everyone conforms to \( R \);
2. almost everyone expects almost everyone else to conform to \( R \);
3. almost everyone has approximately the same preferences regarding all possible combinations of actions;
4. almost everyone prefers that any one more conform to \( R \), on condition that almost everyone conform to \( R \);
5. almost everyone would prefer that any one more conform to \( R' \), on condition that almost everyone conform to \( R' \),

where \( R' \) is some possible regularity in the behavior of members of \( P \) in \( S \), such that almost no one in almost any instance of \( S \) among members of \( P \) could conform both to \( R' \) and to \( R \).

If anyone complains that our final definition of convention is imprecise, he is welcome to use the following quantitative definition.

A regularity \( R \) in the behavior of members of a population \( P \) when they are agents in a recurrent situation \( S \) is a convention...
to at least degrees $d_0$, $d_1$, $d_2$, $d_3$, $d_4$, $d_5$ if and only if it is true that, and it is common knowledge in $P$ that, in a fraction of at least $d_0$ of all instances of $S$ among members of $P$,

(1) everyone in a fraction of at least $d_1$ of all those involved conforms to $R$;
(2) everyone in a fraction of at least $d_2$ of all those involved expects everyone else in a fraction of at least $d_1$ of all those involved to conform to $R$;
(3) everyone in a fraction of at least $d_3$ of all those involved has approximately the same preferences regarding all possible combinations of actions;
(4) everyone in a fraction of at least $d_4$ of all those involved prefers that any one more conform to $R$, on condition that a fraction of at least $d_1$ of all those involved conform to $R$;
(5) everyone in a fraction of at least $d_4$ of all those involved would prefer that any one more conform to $R'$, on condition that a fraction of at least $d_1$ of all those involved conform to $R'$,

where $R'$ is some possible regularity in the behavior of members of $P$ in $S$, such that for a fraction of at least $d_5$ of all pairs of an instance of $S$ among members of $P$ and an agent involved therein, the agent could not conform both to $R'$ and to $R$.

He may go on to define a convention as any regularity that is a convention to at least certain set degrees, which he may pick however he likes.

Let us define the degree of conventionality of a regularity $R$ as the set of sextuples $\langle d_i \rangle$ such that $R$ is a convention to at least degrees $d_0$, $d_1$, $d_2$, $d_3$, $d_4$, $d_5$. We can compare regularities with respect to their degrees of conventionality: $R_1$ is more conventional than $R_2$ if and only if the degree of conventionality of $R_2$ is a subset of the degree of conventionality of $R_1$. It would be interesting to find a single number that measures the degree of conventionality of a regularity; but all the ways I know to do this seem very artificial. If $R$ is a
convention according to the strict definition at the beginning of this section, then \( R \) is a convention to at least degrees 1, 1, 1, 1, 1, 1, and no other regularity can be more conventional.

5. Consequences of Conventions

Suppose \( R \) is a conventional regularity; and suppose \( R^* \) is some logical consequence of \( R \). Is \( R^* \) therefore a convention in its own right?

There are trivial consequences of conventions, and we are not concerned with these. Let \( R \) be our convention of driving on the right; a logical consequence of \( R \) is that we drive on the surfaces of the roads, not ten feet in the air or ten feet underground. More trivially still, a tautology that is a consequence of anything is a consequence of any convention. What we want to consider are the consequences of conventions which depend on convention. Our consequence \( R^* \) depends on \( R \) only if there is a regularity \( R' \) that is an alternative to \( R \) (in the sense of section 3) and \( \neg R^* \) is a logical consequence of \( R' \).

If so, \( R^* \) may be a convention. Suppose you and I want to meet every week; and suppose we spend alternate weeks in different towns \( T1 \) and \( T2 \). Town \( T1 \) has three acceptable meeting places: \( P11, P12, \) and \( P13 \). Town \( T2 \) also has three acceptable meeting places, each analogous to the like-numbered place in \( T1 \): \( P21, P22, \) and \( P23 \). Our convention \( R \) is this: in the weeks we spend in \( T1 \) we go to \( P11 \), and in the weeks we spend in \( T2 \) we go to \( P21 \). A consequence \( R^* \) of \( R \) is this: in the weeks we spend in \( T1 \) we go to \( P11 \). It is a dependent consequence, since \( \neg R^* \) would be a consequence of most of the alternatives to \( R \). \( R^* \) is certainly a convention. In the situations to which \( R^* \) applies—our weeks in \( T1 \)—it is common knowledge among us that we conform to \( R^* \), we expect each other to conform to \( R^* \), and uniform conformity to \( R^* \) is a coordination equilibrium in a coordination problem. In general, a specialization of a convention is a convention. Perhaps a consequence of a convention is a conven-
tion in its own right only if it is a specialization of the original convention.

Now let us look at a dependent consequence of a convention which is not itself a convention. Suppose there is just one town with three acceptable meeting places: $P1$, $P2$, and $P3$. Suppose we want to meet; but in case we fail to meet, it is desirable that one of us should go to $P3$. Suppose our payoff matrix is as given in Figure 29, and suppose

![Figure 29](image)

our convention $R$ is to go to $P1$. Let $R^*$ be the regularity of going either to $P1$ or to $P2$. $R^*$ is a consequence of $R$; and it is a dependent consequence, since not-$R^*$ would follow from the regularity of going to $P3$, which is an alternative to $R$. We conform to $R^*$ and expect each other to, and both of these facts are common knowledge between us. But $R^*$ is not a convention because, I contend, it is not the case that each of us prefers to conform to $R^*$ conditionally upon the other's conforming to $R^*$. Given only that you will conform to $R^*$, with no indication of whether you will do so by going to $P1$ or by going to $P2$, I prefer to violate $R^*$ by going to $P3$. The same is true for you with respect to me.

The case is not entirely clear, however. Consider that what we call a preference conditional upon some state of affairs $A$ is almost always conditional also upon some background state of affairs $B$, which we
regard as a fixed part of the environment. To say that I prefer to drive on the right if others do is really to say that I prefer to drive on the right if others do and if various familiar facts about the causes and effects of collisions continue to hold. Now it is a fact, and common knowledge between us, that if either of us conforms to \( R^* \), he will do so by conforming to \( R \); in other words, it is a fact and common knowledge that we will not go to \( P_2 \). If this fact were included in the fixed background, then each of us *would* prefer to conform to \( R^* \), conditionally upon the other’s conforming to \( R^* \) and upon background. I am sure it is wrong to include in the fixed background this fact, that if either of us conforms to \( R^* \) it will be by conforming to \( R \). But I have no theory to explain why it is wrong. Roughly, the reason is this: in considering preferences for actions conditionally upon actions, the background ought to be kept neutral as to whether actions of the general sort under consideration are done or not.
References


David Kellogg Lewis (1941–2001) was born in Oberlin, Ohio, of academic parents, earned a B.A. in philosophy at Swarthmore College, and an M.A. and Ph.D. at Harvard; during his postgraduate studies he worked for some time at the Hudson Institute, where he became interested in game theory and its potential for addressing conceptual problems. His Ph.D. dissertation became the book *Convention: A Philosophical Study* (Harvard University Press, 1969), a portion of which is reproduced here by permission. He taught at UCLA from 1966 to 1970 and then spent the remainder of his career at Princeton, also spending much time in Australia. His other writings include *Counterfactuals* (1973), *On the Plurality of Worlds* (1986), and *Parts of Classes* (1991). His numerous papers have been collected into five volumes.