Lewis's Notion of a Convention

KEITH COLEMAN
University of Kansas

It has often been claimed, even by philosophers as early as Aristotle,[1] that language use is conventional. What words or expressions have come to be about or refer to has been established by a long tradition of their use in speech acts so that linguistic meaning must be viewed in terms of a common, shared behavior pattern reinforced by what appears to be a tacit agreement among members of a speech community. In his early paper "On Referring", P.F. Strawson[2] claimed that there were two sorts of conventions involved in language use: since meaningful expressions are used to refer to an object or attribute a property to an object referred to, there are conventions for referring and conventions for ascribing or attributing. The meaning of an utterance was then secured by these conventions together with the context of utterance. The importance of the role that conventions play in the generation of meaning was something that Strawson at that time considered to have been largely ignored by logicians. More recently, W.V. Quine and others have challenged this idea and have pointed out a basic implausibility: if the utterances comprising a language are endowed with a meaning on the basis of an agreement or convention among the language users, then it would appear that there must have been a convening of the people who took part in the original agreement and, if circularity problems are to be avoided, these original language users must have, somehow, reached an agreement without the aid of any language. David Lewis in his 1969 book, *Conventions*[3], fully acknowledges Quine's reservations but attempts to rescue the notion of a convention from the doubts of such skeptics by showing how a certain state of affairs can give rise to mutual expectations among the individuals of a population in such a fashion that a regularity in their behavior is reinforced and, consequently, perpetuated. This regularity in behavior, which occurs among a group of people who share a common interest in coordinating their activities, is what Lewis comes to mean by a 'convention'.

---

[1] "On Propositions", 16a20-29, and 17a1-6 in *Aristotle: Selected Works*. trans. Hippocrates G. Apostle and Lloyd P. Gerson. Grinnel: Peripatetic Press, 1983.

[2] Stawson, P.F., "On Referring", reprinted in *The Philosophy of Language*. ed. A.P. Martinich. New York: Oxford University Press, 1985, pp. 220-235.

[3] Lewis, David. *Conventions: A Philosophical Study*. Cambridge, Massachusetts: Harvard University Press, 1969.

Lewis begins his exposition by presenting a series of eleven examples of the type of situation, what he calls a 'coordination problem', in which a convention may arise. Suppose, to take one of his examples, that you and I desire to meet at a certain time tomorrow. Let us say that the particular place where we meet is a matter of little consequence to us both just as long as we both get there at the same time. Regardless of where we both go, if we fail to meet, we are both equally disappointed. In this situation, you and I both try to go to the same place, and for each of us where we finally go is dependent on our expectations about where the other person will go, and this expectation is in turn based on where the other person expects us to go. In other words, I will go to wherever I expect you to go, and where I expect you to go depends upon where I think you expect me to go. If we are fortunate and meet at some place, we may continue this practice if the need to get together persists, and our mutual expectations about each other's expectations are strengthened after every successful meeting.

The important feature of this example that it shares with Lewis's other examples defines the type of problem a group of individuals may face which Lewis wants to single out as being responsible for the genesis of conventions. All problems of this type involve at least two people who are confronted with a situation involving interdependent decision in which there is more than one possible standard of behavior such that uniform conformity to the standard is equally beneficial to all whereas any degree of non-conformity to the standard is equally disadvantageous to all. This type of problem in which agents try to coordinate their activities by conforming to some mutually beneficial standard or pattern of action is called by Lewis a coordination problem, and, by employing some of the concepts of game theory, Lewis is able to give a more precise definition of a coordination problem, a definition which, however, he must necessarily modify in the subsequent course of his analysis.

In the theory of n-person games, agents are construed as acting in such a fashion that the benefits (or payoff) to any one agent may depend upon, according to the situation, not only the agent's action but the actions of all the other agents as well. In order to present the array of information in a systematic fashion, game theorists construct a n-dimensional matrix (where n is the number of 'players' in the game) with dimensions p_1X $p_2X....Xp_n$ where each p_i is the number of possible actions available to the i^{th} agent. Each bounded figure in the matrix (whether a square, cube, or whatever) contains a sequence of n numbers which gives the payoffs to the n agents. Lewis considers, for the most part, only two person games and hence constructs mostly two-dimensional matrices consisting of p_1 rows and p_2 columns. Any game that can be represented by a payoff matrix is of a type that can be considered to be on a continuum between two basic types of game. Games of pure conflict "can be represented by a payoff matrix in which the agent's payoffs (perhaps after suitable linear

rescaling) sum to zero in every square."[4] Games of pure coordination, on the other hand, can be represented by a payoff matrix in which the agent's payoffs (again, perhaps after suitable linear rescaling[5]) are the same in every square. Games of pure conflict are games where the interests of the agents totally conflict, and games of pure coordination are games where the interests of the agents perfectly coincide. Most games theorists actually encounter are neither of these two types but contain elements of both conflict and coordination. The matrix in figure 1 represents a game of pure conflict, while the matrix of figure 2 represents a game of pure coordination. Figure 3 gives a matrix that represents a game involving both conflict and coordination.

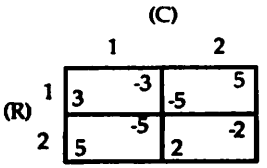

Figure 1

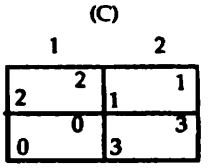

Figure 2

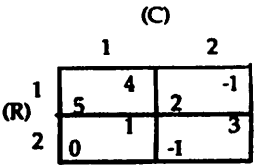

Figure 3

'R' and 'C' designate row chooser and column chooser, respectively, and the lower left number in each square gives R's payoff while the upper right number is C's payoff. Each square, then, gives the payoffs to the two agents when R chooses the action represented by the row and C chooses the action represented by the column. In figure 1, each pair of numbers in the matrix sums to zero, and in the matrix of figure 2 each pair of numbers can undergo appropriate rescaling so that the payoffs are the same in every square.

Lewis is interested in the games of coordination but is not, at least initially, concerned about excluding from consideration games that involve elements of conflict as well as of coordination. His analysis depends on the identification of a class of games in which coincidence of interests predominates. In order to specify this class, Lewis needs to define some other basic game theoretical terms. An equilibrium is a combination of actions (a square in a 2X2 matrix) such that neither individual, acting alone, could have, by choosing a different course of action, improved his outcome (increased his payoff). A coordination equilibrium is a equilibrium in which neither agent could have improved his outcome regardless of how either one, but not both, of them had acted

---

[4]Ibid, pg. 13.

[5]This rescaling is accomplished, square by square, by multiplying the payoffs by some constant and then adding some constant to the products.

differently. Finally, a proper coordination equilibrium is a coordination equilibrium with an outcome both agents prefer to any outcome any one of them could have achieved by either one, but not both of them, acting differently. (The combination (i,j) is an equilibrium iff for any x and for any y, R's payoff at (i,j) > R's payoff at (x,j) and C's payoff at (i,j) > C's payoff at (i,y). The equilibrium (i,j) is a coordination equilibrium iff for any x and for any y, R's payoff at (i,j) > R's payoffs at both (x,j) and (i,y) and C's payoff at (i,j) > C's payoffs at both (x,j) and (i,y). The coordination equilibrium (i,j) is a proper coordination equilibrium iff for any x and for any y (not x=i and not y=j), both R's payoff at (i,j) > R's payoffs at both (x,j) and (i,y) and C's payoff at (i,j) > C's payoffs at both (x,j) and (i,y).

The type of game with a significant amount of coincidence of interest which Lewis is concerned to single out can now be defined. A coordination problem is a game that has a payoff matrix with more than one proper coordination equilibrium. The matrix in figure 2 represents a coordination problem since (1,1) and (2,2) are proper coordination equilibria, but the matrix in figure 3 does not represent a coordination problem since it contains but a single proper coordination equilibrium at (1,1). This more precise definition captures the essence of what Lewis had less formally recognized in the kind of problem typified by his eleven examples. Proper coordination equilibria are states of affairs both agents strive to achieve since they are situations that are mutually beneficial; all other options available to the agents are less than desirable for all concerned. Since there are at least two proper coordination equilibria in any coordination problem, the agents can coordinate their activities in a mutually beneficial fashion in more than one way. There are, however, a few features of the former definition which are not carried over into the formal definition. Lewis amends his definition later on in the text by considering a few important restrictions which preserve these features.

Lewis describes the ways in which a group of individuals confronted with a coordination problem may in practice come to solve it (i.e., reach a coordination of their activities that is satisfactory to all). Sometimes it happens that people just hit upon a happy combination of actions by mere luck, without any conscious deliberation by anyone concerning anybody's expectations. Most of the time, however, people confronted with a coordination problem cannot rely on such a fortuitous happenstance and are forced into a deliberative situation in which each agent must decide what to do based upon what he expects the others will do. In those coordination problems characterized by the occurrence of a unique proper coordination equilibrium that is preferred by all agents to any other proper coordination equilibrium, the coincidence of the agent's actions is guaranteed so long as a knowledge of the situation is common to all participants and everyone is both rational and looks out for at least their own interests and expects everyone else to be both rational and concerned with at least their own interests. This type of problem is, for

Lewis, an example of a trivial coordination problem, and its solution does not exemplify the strategy which Lewis proposes as the way in which coordination problems are in general solved.

When two or more agents are faced with a coordination problem, each of them attempts to come to some reasonable assessment of what to do based upon what he considers to be the most likely action the other one will take. It must be assumed by everyone involved that all agents are rational, share a common background (at least to some extent), and have the same inductive standards. All individuals must have knowledge of what the options are and be willing to bring their actions into conformity with a pattern the conforming to which is a coordination equilibrium on the condition that all others will do likewise. Under these conditions, each agent will try to develop expectations about the actions others will take using whatever information is available. If one has good reason to believe that certain courses of action are likely (or unlikely) to be taken by a certain individual(s) based upon knowledge of, say, preferences or past behavior, then one's expectations may be quite quite strong; in the absence of such information, on the other hand, they may be quite weak. Each agent then decides whether or not to conform to some particular mutually recognizable pattern of action depending upon which action will maximize his expected value. Given that the cost of nonconformity is outweighed by the benefits of achieving conformity, then the rational agent will opt to conform if he views the likelihood of conformity on the part of others to be greater than the likelihood of their nonconformity; if this latter condition is not the case, he will choose not to conform.

If agents A and B are deliberating about their future course of action, then each will develop expectations about what the other one will do. What the other one will do is, however, dependent upon what he expects the first one to expect him to do. This expectation in turn leads to more and more complicated expectations involving the expectations of both agents. A comes to expect that B will perform some action p, and B in turn expects that A will perform some action q. These are what Lewis calls first-order expectations. From these first order expectations together with the assumptions they make about each other's rationality, background information, desires, etc., A and B come to have what Lewis calls higher-order expectations. A expects that B will come to expect that A will perform q, and B expects that A will come to expect that B will perform p. A then expects B to expect that A will come to expect that B will perform p, and B also expects A to expect that B will come to expect that A will perform q. These expectations may be present in the agent's deliberations up to about the fourth order, although in principle there is no upper bound. Neither A nor B have the same first or higher order expectations, but there is a sense in which the higher the order of the expectations which A and B have the greater is their common understanding and agreement of the facts in the matter. A will then come to act in a manner

consistent with both what he expects B to do and what he expects B to expect that he will do, and B will come to act in an analogous fashion. The more information about one another which is shared by the agents in a coordination problem, the higher the order of expectations that will develop and, consequently, the greater will be their chance of solving their coordination problem.

Lewis mentions[6] three factors which when present in the state of affairs involving agents in a coordination problem are particularly effective in producing a 'system of mutual concordent expectations' among those agents. If all the participants take part in an agreement, then, provided everyone has good reason to believe that everyone else will abide by the agreement, each of them will develop strong expectations regarding the other's future actions. If somehow, perhaps by chance, a group has solved their past coordination problem by achieving conformity to some regularity, then this situation may have set a precedent in which there is now a tendency, which is mutually recognized among all participants, for everyone to attempt to solve similar coordination problems in the future by achieving conformity to the same, or similar, regularity. Agents may also come to have strong mutual expectations by recognizing (and realizing that other recognize, etc.) one possible regularity as being particularly salient due to its 'naturalness', convenience, or whatever.

What is it that is common to states of affairs in which a system of mutual concordant expectations is produced and that can account for their production? Lewis contends that there are three conditions any state of affairs A must meet in order for it to generate higher order expectations. If what is meant by a sentence of the form 'A indicates to x that p' is that if x should have reason to believe that A holds then x would thereby have reason to believe that p, then the three conditions can be stated as follows (where P is any population and p is any proposition or propositional content).

(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 (p).[7]

If these three conditions are satisfied by some A then, for a particular p, it may be said that it is common knowledge in P that p. An A meeting all the conditions is then sufficient to satisfy among the members of P any higher order expectation about p given that the members of P make mutual assumptions regarding each other's rationality, inductive standards, and

---

[6] *Convention*, pp. 33-41.
[7] Ibid, pg. 56.

so forth. A provides the particular content for the belief in p among the members of P,[8] and that's its role in the production of the complex of expectations. States of affairs characterized by the factors of agreement, precedence, or salience often provide that content and hence satisfy conditions (1)-(3).

With the established notions of a coordination problem, a coordination equilibrium and common knowledge, Lewis advances the following definition of a convention.[9]

> 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 is common knowledge in P that, in any instance of S among members of P,
>
> (1) everyone conforms to R;
>
> (2) everyone expects everyone 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.[10]

Lewis notes that agents may be involved in a convention and not realize that it is a convention inasmuch as 1, the knowledge that the regularity has features which make it fit the definition may never be realized but remain 'potential' knowledge, and 2, the knowledge about the regularity or the recurrent situation may not be verbal but be derived from immediate empirical assessments of the situation, and 3, the knowledge the agent has may be limited to the behavior of individuals taken one at a time within a particular instance of the recurrent situation without the associated recognition of a general pattern of conformity to a regularity.

However, Lewis finds it necessary to modify the above definition of a convention. He does not wish to rule out recurrent situations that consist of a series of interrelated problems of interdependent decision none of which considered by themselves is a true coordination problem. For example, each competing merchant in a community may set prices in the short run so as to maximize his own profits at the expense of the others, but, since all price changes are always soon reflected in the prices set by all the merchants, none of the merchants stands to gain in the long run,

---

[8]Or as Lewis puts it, A is the basis in P for common knowledge that p.

[9]This is actually Lewis's second, although his first more refined, formulation of the definition of a convention. Lewis abandons his first definition in favor of the following account since an essential feature of a convention was left out: a conventional regularity must be the result of a system of mutually concordant higher order expectations produced as a result of a basis for common knowledge in the population of agents.

[10]*Convention*, pg. 58.

and, consequently, the merchants in the long run collectively either succeed in setting prices at adequate levels or fail to do so and go out of business. Each interdependent pricing decision is not when considered by itself in the short run a coordination problem, but the entire series of such pricing decisions made by all the merchants over an appropriate period of time does constitute a coordination problem. As long as the series consists of problems in which coordination of interests predominates and the whole series when considered together is a coordination problem, the recurrent situation, in this case a series of recurrent interrelated events, is an appropriate S in the above definition.

In order to avoid a difficulty, an example of which will be discussed later, a restriction, though, must be placed upon S, regardles of whether S is a self contained coordination problem or is a series of situations only the whole of which is a coordination problem, in the form of a requirement that agent's payoffs for each particular combination of actions be the same or nearly so so that the coordination problem is at least approximately a game of pure coordination. What this in turn means is that each agent prefers that everyone conform to R given that at least all but one (whether himself or someone else) conforms to R. Since associated with any actual convention there must be at least one other possible convention which could have been established (in order for the actual convention to be in a sense arbitrary), S must also be a coordination problem that contains a least one other coordination equilibrium which corresponds to another possible convention R' such that R and R' are incompatible (in the sense that conforming to both R and R' is impossible). The other possible convention R' must, of course, satisfy all the conditions placed upon R except (1) and (2).

In the final modification, Lewis amends his definition of a convention to allow for cases in which not all the members of P conform to the regularity in every instance of S. The final, less quantitative, definition of the two that he presents near the end of the second chapter is as follows.

> 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 preference 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.[11]

What Lewis endeavors to accomplish with the modifications to his original definition I think can be seen more clearly when depicted in terms of S's payoff matrix. By requiring that, directly or indirectly, S be a game of pure coordination, that S contain another coordination equilibrium which could have given rise to a different convention, and that each agent prefers that everyone conform to R given that..., Lewis is in effect stipulating that for any R which may be a convention R corresponds to some coordination equilibrium on a standard form matrix that is a payoff matrix for S.[12] By a 'standard form matrix', I refer to a matrix, such as the one in figure 4, in which 1, the payoffs in each particular square are the same, either all zeros or all ones and 2, there is a series of two or more squares that contain non-zero payoffs along one of the matrix's main diagonals.

		(C) 1	(C) 2	(C) 3
(R)	1	1, 1	0, 0	0, 0
	2	0, 0	0, 0	0, 0
	3	0, 0	0, 0	1, 1

Figure 4

The squares with non-zero payoffs represent combinations of actions which are in conformity to some possible convention R, and the squares with zero payoffs represent combinations of actions in which either one or both of the agents is not in conformity with any possible convention R. No square of the matrix represents any combination of actions in conformity with more than one possible convention R. In the matrix of figure 4, (1,1) is

---

[11] Ibid, pg. 78.

[12] S's payoff matrix may be one that can be put into standard form by linear rescaling. This rescaling is accomplished by multiplying all the payoffs in every square by some constant and then addding some constant to all the products.

a combination of actions in conformity with a possible convention R_1, and (3,3) is a combination of actions in conformity with a possible convention R_2;all other combinations of actions are in conformity with neither R_1 nor R_2.

In order to see why S must be a game that is at least very similar to a game of pure coordination, consider the matrix of figure 5.

		(C) 1	(C) 2	(C) 3
	1	6, 6	0, 5	0, 0
(R)	2	5, 0	5, 19	6, 20
	3	0, 0	20, 6	5, 5

Figure 5

(2,3) and (1,1) are proper coordination equilibria, so the matrix is a payoff matrix for a coordination problem. Suppose that R-chooser has convinced C-chooser that R is going to do 2 and that C believes this and has good reason to believe it given R's apparent integrity and the fact that the problem has been solved previously by R's doing 2. Suppose R believes that C is most likely going to do 3;R certainly has good reason to believe this given C's payoff at (2,3) and R's belief that C believes R intends to do 2. Higher order expectations could very well be produced in both R and C, and yet R has deceived C in leading him to believe that R intends to do 2. R then reasons that, since C is going to act based upon his belief in what R intends to do, C is most likely to do either 2 or 3 and since the difference in C's payoffs between (2,2) and (2,3) is slight and C will want to avoid the zero outcomes at (1,3) and (3,1) C may well opt to do 2, instead of three, and thereby guarantee that his outcome will be 5 or greater. If he does, R's payoff will be much improved if R does three since R's payoffs at (2,2) and (2,3,) are much less than they are at (3,2). R then chooses to do 3, and, as a result, regardless of how C chooses, a non-coordination equilibrium is reached. The trouble with the situation represented by the payoff matrix is that although it is a coordination problem it contains a degree of conflict: C's best outcome occurs at (2,3), a coordination equilibrium, while R's best outcome occurs at (3,2), not a coordination equilibrium. This type of problem is ruled out as a possible recurrent situation S by the requirement that in every square in the payoff matrix the payoffs for each agent be nearly the same.

There are in Lewis's analysis of convention a few considerations he has not adequately taken into account. Firstly, in cases where a conformity

to a regularity is not initiated by its salience, how is it possible for a precedent to be set? (An agreement cannot really be said to initiate a conventional regularity since an agreement is, by itself, no assurance of future conformity inasmuch as it represents no guarantee that the agents involved have any better knowledge of what each other is likely to do unless it is connected with past behavior which sets a precedent.) If the agents just happen to solve the initial coordination problem by sheer luck, then they will all probably realize it was a coincidence and will not in general display a tendency to repeat the behavior. Secondly, when, and under what conditions, may it be said of a population that its members are no longer party to a convention? In principle, an instance of the recurrent situation S may not be present to any of the members of P at some time(s) and yet the convention has not been violated at any time and remains in effect. A group of jurors is still party to the convention of language even though there are times when its members must remain silent. For Lewis, a convention, once established, is self perpetuating in that past conformity to R leads to current expectations among members of P that conformity to R will continue in the future which in turn ensures their present conformity to R which then will serve as a basis for their future expectations that conformity to R will be maintained which in turn ensures their future conformity to R. How then do the members of a population faced with the continuing need to coordinate their actions succeed in discontinuing a prior convention and adopting a new one if a convention, once established, is self perpetuating as long as the desire for coordination persists? Thirdly, is it really true to say that conventions are maintained by a self-perpetuating system of mutual concordent expectations? In other words, just how seriously do we need to take Lewis's claim that conventional regularities are established and maintained by such expectations on the part of the parties involved? Must all, or most, of the agents literally have these expectations, or must it only be the case that such a pattern of expectations could in principle be produced by the deliberations of all the agents? It would seem, in some cases at least, that when a conventional regularity was first established the participants did have expectations concerning everyone's future conformity, since it was pertinent at that time for everyone to give consideration to the issue of his conformity, but that after the regularity was well established people were conditioned to conform and responded out of mere habit. It seems, for instance, that among experienced American drivers who are sober the issue of their driving on the right side of the road is not an issue they give even a second thought to and continue the practice by force of habit. On Lewis's account, then, driving on the right side is not a convention since it is not reinforced by any reasoning on the part of the agents involved, but this exclusion seems both arbitrary and counter-intuitive.

There may also be a more fundamental difficulty in Lewis's account of the origins of conventional regularities. Exactly how does precedence

work to develop the mutual expectations of the agents? The factors of agreement and precedence would appear to be insignificant unless each agent involved in a coordination problem presupposed of one another that a behavior pattern once displayed would be repeated by each agent in the future. The particular behavior pattern would have allowed the agents to solve the coordination problem in the past and conforming to it in the future would be the very convention that supposedly is just now being established. Lewis's account here seems to beg the question in that his explanation requires the explanandum in the explanans.

Lewis's formal treatment of the notion of a convention has succeeded in making explicit the general idea that conventions, on the one hand, are in an important sense arbitrary and yet, on the other hand, the continuation of a conventional practice is essential to the well being of all involved. It is still a question, however, whether or not Lewis's notion of a convention, and of a conventional signaling system, can account for linguistic meaning. Donald Davidson has argued[13] that such an account of convention will only explain in what sense language *use* among members of a community is conventionally governed and will not account for language meaning. The members of a speech community, due to their participation in a linguistic convention, will expect (and expect that others will expect...) that expressions with the same or similar meaning will be uttered under similar circumstances to bring about the same or similar intended effect in an audience.

[13]See Davidson's essay, "Communication and Convention," in *Inquiries into Truth and Interpretation* (Oxford: Oxford Univ. Press, 1985).