JEAN-PIERRE DUPUY

COMMON KNOWLEDGE, COMMON SENSE

1. SPECULARITY AND COMMON KNOWLEDGE

This essay is dedicated to Franco-American friendship. Unfortunately, intellectual relations between these two countries are impaired by the fact that we frequently pay no attention to each other's philosophies, and at times, members of the two traditions bear an attitude of reciprocal contempt. The French "sciences of man" (be they structuralist, poststructuralist, or deconstructionist), do, of course, have a large following throughout North America – a following which is today even greater than in France. Yet in most American universities, these currents of thought remain confined to the literary disciplines and are rarely taken seriously in philosophy departments. No dialogue, not even a polemical one, takes place.[1]

I would like to suggest that the concepts of *play* and *game*, or in French, *le jeu*, could provide the occasion for some fruitful exchanges between French and American philosophers. As is well known in many American circles, these notions are important to such thinkers as Lévi-Strauss, Lacan, and Derrida, yet French thinkers are less familiar with the ways in which "analytic" philosophers have related to them. The notion of "common knowledge," which will be at the center of the present discussion, first made its appearance in the "philosophy of mind" (an Anglo-American specialty), and was initially conceptualized within the framework of formal game theory. It strikes me as being obvious that there are many connections between this vein of work and some of the most typical themes in contemporary French philosophy. Yet these connections must be made explicit, which implies that we should pay at least as much attention to the differences as we do to the similarities.

Let us begin with the situation in France. Lacan, for one, saw game theory as the ideal framework for demonstrating the central thesis of his version of structuralism, namely, the idea that "the unconscious is the discourse of the Other." This central thesis had its variations and

Theory and Decision **27** (1989) 37–62.

corollaries. For example, there was the notion of the "autonomy of the symbolic order," the idea that "the symbolic is the world of the machine," and again, the emphasis on the supremacy or transcendence of the symbolic over the imaginary.

How are the Lacanian categories of the "symbolic" and the "imaginary" related to formal game theory? First of all, it may be noted that in game theory the very rationality of the players implies that they must put themselves in each other's shoes so as to examine the situation from the adversary's viewpoint. In so doing, each player perceives that the other has done the same in regard to him. The result is a play of mirrors, a specularity that is potentially infinite. This is for Lacan the realm of the imaginary with its phenomena of mimesis or imitation, identification, and conflict. Second, game theory shows that when individuals thought to be rational and free are allowed to interact within a given structure, a world of necessary laws emerges. It would seem that these "free" individuals were in fact subjected to the law of an "Other" – which is precisely what Lacan has in mind when he speaks of the symbolic. Many literary critics are familiar with Lacan's use of these categories in his famous "Seminar on the Purloined Letter," which includes an analysis of the game of odds and evens described by Edgar Allan Poe. What is at stake in this analysis is nothing less than a "deconstruction" of Freud's concept of the "death drive." Interpreted in terms of repetition automatism (*Wiederholungszwang*), this concept receives a formal description in the light of game theory and the theory of automata. We should also mention Lacan's analysis of the game of the three prisoners in another essay in his *Ecrits*, "Logical Time and the Assertion of Anticipated Certainty," for we will be encountering a variation on this game in what follows.

Approaching Common Knowledge

First Example: Peggy's hair dryer is broken. She would like John to fix it, but does not want to ask him to do it. So she imagines the following strategy. She takes the machine apart and leaves the pieces lying around on the table, as if she had been trying to fix it herself. But she also arranges it all so that John can understand that she was only pretending to fix the machine. Her real intention is to inform John that

she wants his help and that all this is mere pretense, and the way she finds of conveying this information is to make it manifest to John that she has the intention of informing him of this. Yet this second-level intention (the intention to make it manifest that she has an intent to inform him) must remain hidden from John.

This last condition specifies what makes Peggy's strategy different from a situation of "overt" communication, that is, one in which Peggy would simply ask John directly for his help. But before we undertake a more precise analysis of this difference, let us see what its implications are for the interactions between Peggy and John. These implications are quite important. Peggy does not want to risk John's refusal of her request, nor does she want to owe him anything. But in making a direct request, she will necessarily confront one of these unseemly alternatives. Her strategy of pretense, on the other hand, allows her to avoid both of them: if John works on the hair dryer, he does so of his own free will; having requested nothing, Peggy owes nothing. But John does not have to work on the machine, for he (believes that he) is not supposed to have interpreted Peggy's actions as a request for help. Should he not help her, it will not be a matter of a painful refusal, but of a simple lack of attention on his part.[2]

Let us note, then, the following proposition, *P*:

Peggy's intention is to ask for John's help.

Given Peggy's behavior, we can allow that the following propositions are true:

(a) Peggy "knows" *P*;

(b) John "knows" *P*;

(c) Peggy "knows" that John "knows" *P* (she has arranged it this way). In other words, Peggy "knows" (b);

(d) John does not "know" that Peggy "knows" that John "knows" *P* (if Peggy has realized her goal, which was to make him believe that her pretense was innocent, John believes that Peggy does not "know" that he "knows" *P*). In other words, John does not "know" (c).

Certain authors claim that in a situation where Peggy asked John overtly for his help, it would be necessary to hold as true the *infinity* of the following propositions:

(a) Peggy "knows" *P*;

(b) John "knows" *P*;
(c) Peggy "knows" that John "knows" *P*;
(d) John "knows" that Peggy "knows" *P* etc., to infinity.
By definition, this is the same as saying:

> *P* is "common knowledge" (henceforth: CK) between Peggy and John.

The main criticism directed against this notion is that it involves an actual infinity of steps, collapsing them into a single movement. The human mind, a finite machine, is held to be incapable of reaching this actual infinity. Should it be a matter of deciding whether a given proposition, *P*, is CK among two or more individuals, what would be at stake? An actually infinite number of tests would be necessary, and undecidability should be acknowledged. Schiffer, for example, wants to define the act of communicating in terms of the fact of making one's intention to inform the other a matter of CK; yet this runs the risk of making it impossible to decide, in any given situation, whether communication has taken place.

From David Lewis, the inventor of CK, to Sperber and Wilson, who want to replace it with the idea of "mutual manifestness," only one approach has been adopted in trying to avoid the problem of infinity just noted.[3] That approach amounts to weakening the notion of "knowledge" so as to reduce the *actual* infinity required by CK to a merely potential infinity. (This is why we have put the verb "knows" in scare quotes in formulating our propositions above.) Thus Lewis is concerned with a potential or tacit form of knowledge. Following this kind of knowing, you "know" that Noam Chomsky never had lunch with Shakespeare, even if the very idea never crossed your mind before you read these lines. Sperber and Wilson want to replace the expression "I know that *P*" by the following: "*P* is manifest to me." What is meant by that is that *P* can be perceived or inferred by me. So then we can say that it is "manifest" to you that Chomsky and Ronald Reagan do not play golf together every Sunday afternoon. We should note that in these terms, something totally false can be "manifest" to someone, which is not the case with knowledge. In either case, it is not necessary *effectively* to run through the infinite set of conditions that must be satisfied in order to be sure that a proposition *P* is CK (or "mutually manifest").

In what follows we will be proposing a different kind of solution to the problem of infinity associated with the concept of common knowledge. But it is first necessary to see why the question arises. The communicational and strategic situations that we want to analyze involve specular phenomena. By "specular" is meant the capacity of the human mind to put itself in the place of another and "see" the world from this other party's point of view. As was noted above, once specularity gets started, it can generate a potentially infinite series of the following kind: "I think that you think that he thinks that I think . . ." I will call the degree or level of specularity the number of successive steps of this sort, minus one. CK corresponds to an infinite specularity. But why has this notion been set forth at all? Because there seem to be many communicational or strategic situations where the solutions, seen as functions of the degree of specularity, change abruptly when this degree passes to the actual infinite. Remarkable properties may appear along with CK, properties that could not have been obtained with any form of finite specularity, no matter how great its degree. There is, then, a discontinuity when we reach the infinite case. For this reason, something is actually at stake in asking whether CK is a concept that can be reconciled with the finitude of the human mind.

The three examples that follow illustrate this difference between a finite or null specularity and CK. Let us note in passing the obvious but important fact that "shared knowledge" (everyone knows P) *is not* to be confused with CK. The latter requires much more – indeed, infinitely more!

Second Example: In his study of the forms of economic exchange characteristic of traditional societies, Pierre Bourdieu admits that the structuralist view captures part of the truth: the traditional patterns of exchange apparently take the form of a gift-giving relationship, while their "objective truth" is the law of reciprocity, the sordid law of interest hidden behind the trappings of generosity and cordiality. Yet in his critique of Lévi-Straussian structuralism, Bourdieu claims that this objective truth is not the *whole* truth. Should, for example, this truth be broadcast in the public space – should it become a matter of common knowledge – then the system of exchange would collapse. The truth of this system must be more complex given that it "could not

exist if it was perceived according to the [structuralist] model." The truth of practice *contains* the structuralist truth in both senses of the word: it carries this truth while also withholding it. Everyone knows the structuralist truth, but it is not CK. Bourdieu describes the case of a Kabyle worker who, in asking to be given money for his participation in a collective task instead of accepting the meal traditionally served at the end of a communal job, was thereby disclosing that this meal was already a form of payment and not a pure gift. In making such a statement, the worker was only "betraying the most poorly and most efficiently kept secret of all, the secret that everyone has the obligation to keep."[4]

The system rests, then, on an "open secret," but it is a secret nonetheless. The seemingly oxymoronic term, "public secret," is not really contradictory because the real secret is that there is no secret.

Third Example: The airline company asks you to arrive at the ticket counter some two hours before the actual takeoff time on an international flight. This rule is pronounced in a totally rigid and unflinching way. Yet in fact everybody knows that this rule can be broken; moreover, they all know that efficiency even requires that it be broken: the optimal situation is when passengers come to the counter in a regular flow between T-2 and T-1, not all at once. Yet it is hard to see how the company could make an announcement to the effect that in saying T-2, what is really meant is T-1, thereby making it CK that this rule was made to be broken.

Fourth Example: For quite a long time in the economic history of France, the effective inflation rate was in fact two points higher every year than the officially predicted rate. Now, a rate in which everyone believes is self-fulfilling. Knowing this, and wanting to reduce the increase of prices, the government made public an officially predicted rate, one reasonably lower than what they really expected. But as time went on, people learned the rule and were thus able to figure out the government's true prediction; once it had become their own belief about the future, they made it come true.

Let us assume, then, the following proposition, P: The real inflation rate will be about two points higher than what the government officially predicts.

In such a situation, the government knows P. If it foresees a given rate y, its official prediction will be x, where $x = y - 2$. Now, the economic agents know that P, so that when the government predicts x, they foresee $x + 2$, and thereby bring $y = x + 2$ about. We may also assume that the government officials know that the agents know P, and they cannot admit that they in fact foresee $y = x + 2$, because then the agents would take y for x and add two more points.

Everyone knows what is going on, but that does not mean that CK can be achieved. The rule implicit in P cannot be made public without giving rise to a runaway phenomenon. Once P is public, to predict x is to predict $x + 2$. To predict $x + 2$ is to predict $x + 4$, and so on. In such a situation, truth cannot do without the help of a certain amount of dissimulation.[5]

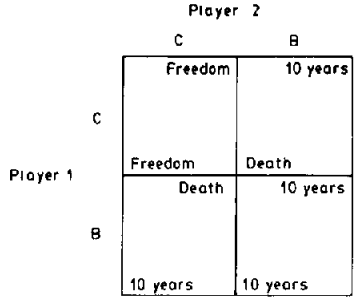
2. COMMON KNOWLEDGE AND COMMON SENSE

Fifth Example: Consider the following game. Two prisoners have been condemned to die and are waiting in their respective cells for the day of their execution. One morning they receive a message from the local dictator informing them that he has decided to pardon them and set them free the very next day. On the other hand, they also have the choice of asking that their punishment be commuted to a sentence of ten years. If they want to choose this option, they must make their request before midnight and it will be granted immediately. Should one of them do so, the other prisoner will be executed, unless he too has requested the "gift" of a ten year-long sentence. And of course the two prisoners are not allowed to communicate with each other about the situation.

Each prisoner must choose between two strategies: do nothing or write to the dictator. The first is a strategy of cooperation (C), and the second one of betrayal (B), because when one prisoner asks for ten years in prison, there is a chance that the other one will be killed. It is not clear what advantage such a betrayal would be to the betrayer, who gets ten years in jail instead of being set free right away. There would seem to be no problem here.

This kind of game is typically analyzed in terms of the following payoff matrix. There are four boxes generated by combining the two

possible strategies with the two players. Each box shows each player's payoff (player 1's payoff is identified in the lower-left corner, and player 2's payoff is in the upper-right corner).



Let us recall a few of the basic concepts of game theory. An *equilibrium* (or Nash equilibrium) is a state of affairs such that each player acts in a way that maximizes his own advantage while supposing the other player's choice to be given. A *coordination* equilibrium is an equilibrium in which each player, whose choice is set, prefers that the other act as he does in the equilibrium in question. If the preference is strict we have a *proper* equilibrium; if not, the equilibrium is improper.

We see that the game described above has two equilibria: boxes CC and BB. The first is a proper coordination equilibrium: if I cooperate, it is crucial to me that the other prisoner cooperates as well. If this obtains, I am free, and if not, then I will be executed. The second equilibrium is an improper coordination equilibrium: if I ask for ten years in prison, I will get ten years no matter what the other prisoner does. So this is a strategy that frees me from having to depend on the other's behavior.

The cooperative equilibrium clearly "dominates" the equilibrium reached through betrayal insofar as it is more preferable for both players. Common sense would have cooperation prevail over betrayal, and one hardly sees why this should not turn out to be the case. Yet it suffices to have the players play the game to see that it is not so simple. As soon as specularity gets under way, it takes the form of suspicion and irremediably upsets the cooperation equilibrium. Only the equilibrium achieved through betrayal stands outside of its pernicious reach.

Two kinds of rationality can come into play in such a game. The first is the *utilitarian* rationality, the kind that is already implicit in the definitions of equilibrium that are given: working within certain constraints, each player tries to maximize his own advantage. Since a very important loss is possible in this situation (death), it is impossible to eliminate a *prudential* kind of rationality from coming into play. Each player can try to *mini*mize the *maxi*mal loss that he may suffer. This is the so-called *minimax* strategy. If I choose to cooperate, I risk death; if I betray the other prisoner, I will get ten years, no matter what happens. Prudence, then, would have me choose the betrayal, even though it is absurdly destructive.

I will ask the dictator for a ten year sentence if:

(a) I am behaving prudently, which is absurdly destructive in this case;

or:

(b) (a) is not the case, but I suspect the other prisoner of behaving prudently, which is absurdly destructive in this case;

or:

(c) neither (a) nor (b) is the case, but I suspect that the other suspects me of behaving prudently, which is absurdly destructive in this case. And so on, to infinity.

The strategy of prudence wins out as soon as suspicion crops up at any finite level, even if it is at a very deep one. It is as if the play of specularity were capable of dredging up to the surface any element of suspicion buried in the depths of consciousness. A player will adopt a strategy of trust only if suspicion does not appear at *any* level whatsoever. Thus an actual infinity of successful tests is required for cooperation to be realized. This infinity can be expressed in a single sentence: the fact that both players set aside prudent rationality because it is absurdly destructive in this case must be CK.

We must concede to the critics of CK that it is psychologically implausible that the infinity of conditions entailed by this concept could be satisfied. Cooperative equilibrium, then, is unstable and the players will choose to betray each other. Yet such a conclusion is shocking. Intuitively we feel a need to save cooperation in the name of common sense. And along with cooperation we must save CK. There is a way to do this: replace the infinite specularity of CK with a

complete absence of specularity. Let us deprive our players, then, of any capacity for speculative or specular flights; incapable of simulating the thoughts of others, they will choose to cooperate in a wholly "natural" way. This may well be a way of mutilating or alienating our players, but there is something emancipatory about this alienation. The following examples should make it possible to state more precisely this equivalence between infinite specularity and the absence of specularity.

Sixth Example: Here is the kind of "specular" game to which Lacan was particularly attracted. The action takes place on an island inhabited by around a hundred couples. Some of the husbands are cuckolded by their wives. If a man is in such an infelicitous state, he is the only one living on the island not to know about it. In fact, there are exactly three such husbands. The rigid mores of the island specify that when a man learns of his wife's infidelity, he must break up with her on the same day before midnight. A foreign missionary visits the island and studies these exotic customs. The day of his departure, he makes a little parting speech. He concludes his talk to the assembled population of the island with the following words: "There is at least one cuckolded husband among you." Two uneventful days pass by, and then on the third day, the three cuckolds break up with their respective wives. Now, what has happened?

This game belongs to a well-known category. A recursive reasoning is the most efficient way to express its logic. Suppose that there is only a single cuckolded husband. Looking around at everyone else, he only sees men who would have – if they knew it – every reason to be proud of their wives' faithfulness. But the missionary just stated that there was at least one cuckold on the island (and we assume that his statement will not be put in doubt). It follows that the unfortunate husband must conclude that he is the cuckold, since he knows that everyone else's wives are faithful. So that night he breaks up with his own spouse. (A full account will also include the demonstration that the other husbands upon seeing this conclude that their own wives are faithful.)

Now let us assume that there are two cuckolds on the island. Each of them entertains the following thoughts: Let's assume my wife is

faithful. In that case there is only the one cuckold whose plight is already known to me. Now I put myself in his place and reason the way he would, thereby arriving at the demonstration of the previous paragraph. This fellow should therefore break up with his wife before the next day. I wait to see this happen and observe that it doesn't. (We, from the outside, understand why: each of the two husbands reasons in the same way.) From this I can deduce that my first premiss is false, and thus that my wife is cheating on me. The next day I break up with my wife, and I feel no surprise when I see the other guy do the same.

Through this recursive logic we arrive at the following rule: let us consider a husband who observes n cuckolds around him and waits for the end of the nth day. If nothing happens, he must conclude that his wife has betrayed him and must break up with her the next day. If, on the contrary, the n cuckolds break up with their wives, the husband is guaranteed of his own wife's faithfulness.

As we said, this game is quite well-known, but only recently has the following question been raised in relation to it.[6] We know that what sets the sequence of specular reasonings in motion is the missionary's public statement to the effect that there is at least one cuckold on the island. Without this *utterance* nothing would have happened, and thus it plays a decisive role. Yet the contents of this statement convey no new information to anyone. Everyone already knows that there is at least one cuckold. The happy couples see three cuckolds in the community, and the cuckolds see two of them. In terms of its informational content, then, the missionary's statement is useless. Yet it is also indispensable. How, then, are we to analyze the effect of the statement made by the missionary? This question may be reformulated. Since the statement brings no new information to any of the islanders, and since its utterance is what makes them go from a state of ignorance to a state of knowledge, to *whom* is this statement really addressed?

Consider the following proposition, P: On this island there is at least one cuckolded husband. Now, before the missionary arrived, everyone on the island knew that P. But P was not CK. This can be observed in the case of two cuckolds, John and Peter. John knows that Peter has been cuckolded; then he knows that P. But John does not know that

Peter knows that *P*. For if John is not a cuckold, Peter sees no cases of infidelity and cannot know that *P*.

When the missionary publicly utters *P*, *P* becomes CK immediately. Not only is it simply true that everyone knows *P*, from now on everyone knows that the others know *P*, they know that the others know that they know, and so on.

There is a more economical way of putting what was just said. Aumann has contributed an important finding which we would reformulate in the following way: there exists a subject of CK. In epistemic logic, we can define "knowledge" by introducing an operator that satisfies certain axioms. Such a definition is purely syntactical. To this syntax, however, corresponds a semantic interpretation by means of which we associate a knowing subject with each knowledge operator. Aumann has shown that given *n* elementary knowledge operators, the operator of CK associated with them satisfies the axioms that define knowledge. The operator of CK is thus itself a knowledge operator. A subject must therefore be associated with it, and this is the subject of CK. Following Lasry, we will call this subject the Other (is this Lacan's "big Other"?).

By construction, we have the following property of this Other: *the Other knows that P if and only if P is CK*. Thus when the missionary says *P*, he is informing the Other. Yet this proposition is not wholly pertinent. In fact it is not hard to show that there is a single subject, the Other, the subject of CK, for whom the following holds true: *the Other knows that P if and only if everyone knows that the Other knows that P*.

When the missionary utters *P*, everyone knows as a result that the missionary knows that *P*. Instead of saying that the missionary addresses himself to the Other, it would be better to say that when he speaks, this Other is incarnated in him, or that he speaks in the name of the Other. Everyone on the island can discover the truth about themselves because they all know that the Other knows.

There are, then, two ways of describing the logic of this game. The first way describes individuals who are anxiously trying to know the truth and who get lost in the infinite play of mirrors created by their attempt to imagine what the others know about them, just as the ancients, blind to the nature of their own *daimon*, tried desperately to scrutinize its reflection in the eyes of their companions.

(Two remarks on the last phrase. Adverbs such as "anxiously" and "desperately" are obviously out of place in discussing the game under consideration, the subjects of which are "purely logical subjects," as cold as an automaton – all of which is perfectly in keeping with Lacan's sense of the word "subject." Even so, the logical specularity mimics the anguish felt by someone who, like Oedipus, has to deal with others, and not simply the Other, in trying to know himself. Second, for any realization of this game, the number of cuckolds n is necessarily finite, and thus it is not necessary to get lost in an infinite specularity in order to arrive at a solution. Only a finite number of the propositions made true by CK will in fact be called upon. Yet CK is indispensable in the *general* theory of the game, a theory valid for any n.)

The second way of describing the game replaces the infinite specularity of CK with a total absence of specularity. The players no longer look to see what the others are doing; they no longer anticipate each other's thoughts, and each individual only relates to the Other. More exactly, each individual only relates to the others through the mediation of the Other. The Other is a fixed point, a node of relations, the incarnation of the group that focuses its gazes upon a single point. One could say along with Lasry that this Other is the "symbolic instance." But in saying this we must be aware of the consequences. The fundamental postulate of French structuralism is that the symbolic transcends the imaginary. The symbolic governs the movements or "play" of the imaginary and is in no way affected in turn by the imaginary order. But in the case we have examined, the Other is produced by the specular game. The agents are guided by a reference point that they themselves have caused to emerge. This kind of "tangled hierarchy" cannot be grasped within the terms of French structuralism.

The movement from an infinite to a null specularity may be noted in yet another way. Let us suppose that the agents know the theory of the game they are playing. In fact, this is not an optional hypothesis because the theory in question can only be established if the agents know it. This "bootstrap" is another manifestation of tangled hierarchy.[7] Thus the agents know the rule that we set forth above: if they see n cuckolds in their community, they will wait to the end of the nth day, etc. Far from being a matter of a dizzying specularity, their mental activities are reduced to a mundane calculation.

Yet this evacuation of specularity is too sudden and abrupt. It is really only the consequence of focusing on an artificial game where a form of exteriority clearly dominates the action – this exteriority being that of the rules, the sharp punctuation of time into days, and the role of the missionary.[8] In the next example, this exteriority is removed.

Seventh Example: The game or contest taken up in this example was run several years in a row by a well-known magazine. Shortly before the announcement of the winners of a major literary award, the magazine publishes a list of ten novels. The idea is that the readers should collectively pick the one they prefer. There is clearly an ideological agenda here: a democratic choice is set in opposition to the opinion of the "experts."

But what exactly is the nature of this democratic choice? The actual rules of the game suggest a fascinating definition of democracy: "The winner will be the one who has chosen the novel having received the greatest number of votes." Now, if this were the only rule, there would be a large number of winners, and thus a second clause is added: among the finalists, the real winner will be the contestant who justifies his or her choice in an essay judged to be the best by a certain jury.

This contest is more diabolical than it appears at first glance. To play well, several traps have to be avoided – in fact, there is an infinity of these traps arrayed in a hierarchy of levels. The error at level zero would be to make a choice in function of one's own preferences or opinion – what guarantee is there that this opinion will correspond to the choice of the majority? Somewhat more subtle is the error at level one. This error is a matter of choosing in terms of what one thinks the preference of the majority will be. But this strategy would only be rational if one supposes that the others all fall into the trap at level zero, each making a choice in function of his or her autonomous, individual preference. Now the recursion has begun, and it can be formulated as follows: the error at level n consists of choosing as if the others made the error at level n-1. The average opinion can only hope to discover what the average opinion is by engaging in this kind of infinite regression or endless play of mirrors, the "mise en abyme" celebrated by so many deconstructionist literary critics. The game in question invites us to take up this exercise of reflexivity and social self-reference.

At this point we have a much more economic manner of restating our previous results because we can now summarize the infinity of traps in a single proposition: the error would be to single out an object, the distinctive feature of which is not CK.

Yet this is obviously not how the players in fact play this game. The socio-cultural group that they make up is immersed in a history, a tradition, a particular world and a particular form of common sense. Each individual has an implicit, unformulated and tacit knowledge of this world, and although this knowledge is not explicit, it is constitutive of the individual's social being. This common sense has been collectively created by individuals, but it nonetheless appears to them as if it were an objective reality wholly external to their own making and doing. We could speak in this regard of "alienation," but only if we remember that "alienation" translates both *Entäusserung* and *Entfremdung*. The former denotes the exteriorization of a form of mediation, whereas the latter involves the estrangement and self-division that characterize an "unhappy consciousness."[9]

The "natural" way to play is clearly for each player to consult his or her common sense, making a judgement without engaging in any specular reference to what others might choose. The others are still present in this agent's individual choice, but it is as if their views had been crystallized into objects. Mediation by means of common sense makes it possible to obtain with null specularity what logic thought only an infinite specularity could obtain. Like the Other in the preceding example, common sense guides precisely those who have collectively created it. The equivalence between infinite and null specularity is realized through the figure of tangled hierarchy.

But the situation can be presented in a rather different way. It is possible to argue that this mediation by means of common sense is what satisfies rather than eliminates the strong conditions involved in CK (namely, the requirement that one must not choose an object on the basis of any characteristic that is not CK). It is not very hard to grant that such propositions as $2 + 2 = 4$, or the sun rises in the East, are CK for particular groups of human beings. To the extent that the representations of common sense are, as Pascal said, like a "second nature" and seem wholly self-evident, we can admit that these beliefs are by the same stroke CK.

Here we see that it is useful to distinguish between two forms of CK,

a CK of crisis and a CK of order. The crisis is a crisis of common sense. Lacking external points of reference with which to coordinate their actions, people become fascinated with each other. They stare at each other and get lost in the mirrors of specularity. The "logical" solution to the contest just discussed mimics this crisis of a specularity stretching out to infinity without ever reaching it. In a society that is in order, common sense appears as being always already constituted, objective, and incontestable. By letting themselves be guided by the standards provided by common sense, people automatically satisfy the actual infinity of conditions required by CK, and they do so without any special mental or affective effort. Infinite and null specularity are in this case indistinguishable.

Eighth Example: In this final example, we see how CK can create a situation of indeterminacy or undecidability. Only an (actual) infinite specularity can produce this result.

The play has two roles: father and son. The father often leaves on business trips and won't let his son use his car while he is away. The father can have two stances, confidence (C) and suspicion (S). In the latter case, he has a neighbor keep watch over the son. The son can either obey his father (O), or disobey him (D).

Act One: The father leaves, confident. The son disobeys him. This case will be symbolized by CD.

Act Two: The father happens to learn that his son disobeyed him. But the son does not know that the father knows this. When the father leaves on another business trip, he has his son watched. Unsuspecting, the son continues to disobey. This is SD.

Act Three: The neighbor's son tells his playmate that the father is having him watched. The son disobeys and the father knows it. The son knows that the father knows, but the father does not know that the son knows. The father remains suspicious. The son now obeys. Now we have SO.

Act Four: The father learns that his son knows that he was being watched. (He is no longer surprised by the son's obedience.) The father becomes confident once more and puts a stop to the surveillance. This is CO.

Act Five: The son learns that the father has found out about his

knowledge. The son figures out that the father has regained his confidence. He takes advantage of it and disobeys. We have returned to CD.

And so the play goes on, running endlessly through the cycle of CD, SD, SO, CO, CD. We move one step forward whenever one of the two agents learns something new about what the other knows about him.

Let us now suppose that at the end of Act One, the father comes home earlier than expected. On the way home from the airport in the taxi, he happens to drive past his son who is at the wheel of the family car. Their looks meet for a second. Like two mirrors facing each other, they reflect back and forth infinitely a truth that has become CK: the son has broken the paternal law! What will happen in the next act? Will the father be trusting or suspicious? Will the son obey or not? There is no way to say. The CK has created a radical indeterminacy. Any finite value of the specularity can determine the system's state unambiguously. But if we are convinced that the crossing of the two agents' glances disrupts this determinacy, we must admit that it can also generate an actual infinity.[10] The exchange of glances has the same quality as those properties that we take as being self-evident in nature or in a naturalized social order.

3. COMMON KNOWLEDGE IN SOCIAL AND POLITICAL THOUGHT

The concept of CK comes into many social and political theories in either an explicit or an implicit manner. Since the examples that we have chosen to illustrate this concept take the form of "games," they could appear to be purely anecdotal and artificial. In fact they were chosen because they each illustrate an important theory. In certain cases, the very authors of these theories have taken advantage of the pedagogical value of these examples. But at this point we must reveal our sources. The notion of CK is used in an astounding variety of ways. It is employed in the analysis of the rules and conventions of groups and societies that are not in crisis; it is also employed in studying cases where the attempt is being made to reconstitute stable conventions in a situation of crisis; and finally, it is used in conceptualizing the constitution of a public space. Once our astonishment

has passed, we may note that this variety of uses has nothing to do with any inconsistency in the concept of CK. Instead, it may be that there is not such a huge distance separating the mechanisms of social order and social disorder.

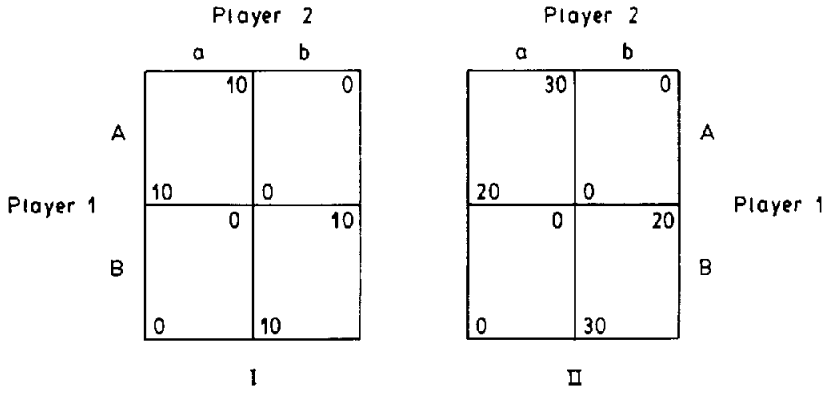
3.1. Social Order and Conventions

As we have already noted, Lewis invented the concept of CK in his book *Convention*. Lewis's concern here was to examine the commonplace idea that there is something "conventional" about the language we use (such as English or French). We could just as well say things differently; the relation between signifiers and signifieds is "arbitrary" or "unmotivated." Yet this intuition clashes with the evident fact that language is hardly the result of some kind of explicit agreement or "social contract" because such things themselves presuppose its existence. In order to save the intuition, it is necessary to make the notion of an implicit agreement or tacit convention more coherent and solid.

Lewis acknowledges two sources or models for his theory. Hume's theory of the origin of justice and private property is a first. In Hume's view, if men allow each other to possess their own goods, this is not the result of a promise or engagement that they have made directly with each other; rather, this results because they are guided by the common sense they share; this same common sense tells them where their interest lies. My interest is not to touch my neighbour's property as long as he behaves the same way towards me in return. There is a coordination of actions in which each individual acts with reference to the actions of others, and in which each individual acts as he does *because* he supposes that others will act the way they do. Such a coordination may be called a convention even though it is not the fruit of any explicit agreement.

Lewis's main source is Thomas C. Schelling's *The Strategy of Conflict*, from which he draws the notion of coordination games.[11] This kind of game has not generated much interest among mathematicians, who deem them too simple. Here are two examples, taking for granted the same conventional form of notation used before.

We are dealing with coordination *situations*: the interests of the



players coincide almost perfectly. Their *problem* is to coordinate their actions. This is a real problem, however, for in game I and in game II, there are two proper coordination *equilibria*: Aa and Bb. The resulting indeterminacy can bring about the disastrous outcomes Ab or Ba if the players do not manage to harmonize their strategies. The problem is even more delicate in game II because the two equilibria give each player in turn a better payoff than the other. But even in this case, the individual loses less by accepting an equilibrium that penalizes him relatively than by trying to have more and thereby taking the chance of there being no coordination at all.

Schelling saw quite clearly that the interest of such games has nothing to do with what one can say about them from a formal point of view, but lies in the cognitive phenomena they generate in practice. For example: a husband gets separated from his wife in a large department store. How do they get back together? Each tries to think of the "obvious" place to meet, but there are several possibilities. The problem is not simply a matter of predicting what the other person will do, because what the other person will do depends on how he or she in turn predicts what the first one is going to do, knowing full well that the latter is also trying to put himself or herself in the place of the other, and so on. A limitless specularity comes into play right away, already prefiguring CK. Schelling shows that in spite of a theoretical indetermination, these problems are quite generally solvable in practice thanks to the cognitive performances of the agents in question, who succeed in getting coordinated *because* they know that the others

are trying to coordinate with them. *So here specularity is a stabilizing factor.* Imagination is more useful than logic. Each party looks for a sign of what the other is likely to think that he is going to think of. This sign is provided by some kind of salient feature or trait that may be striking because of its uniqueness or because of the meaning that it has within their common history or situation. Poetry and humour, symbols and fantasy carry more weight here than mathematical reasoning.

Lewis took up this idea and made use of it in defining his notion of (tacit) convention. A convention is the solution to a coordination problem that, having managed to attract the imagination of the agents, tends to recur. The fact that the coordination problem has several possible solutions and the fact that the one solution that is chosen is the convention in question are held to be CK. Lewis adds that this condition of CK is what guarantees the stability of the convention.[12] The agent's conviction that he should conform to convention is reinforced by his simulation of the reasoning that everyone else performs in deciding to conform to convention.

In this theory the arbitrariness of convention is known to the agents, and is even CK to them, without leading to instability. This would be a case of perfect social transparency if the overall framework were not one where there is a total absence of communication between the agents.[13] As in the economic theory of the marketplace, the agents do not talk to each other and know nothing of promises and commitments. What unifies and totalizes a set of radically separate consciousnesses is CK, with its movement towards infinity.

It is interesting to note that Lewis's theory occupies an intermediary position within the semantic field of the notion of CK, and thus has been criticized from two directions. Some think that the stability of conventions implies that the agents must not know them to be arbitrary.[14] Others criticize Lewis because they believe the theory of conventions should involve even more transparency.[15] These two positions provide a kind of frame for Lewis's theory, and we will now take them up.

3.2. Crisis and Conventions

The magazine contest that we discussed in example seven is essentially

the same as the one Keynes used to illustrate his theory of financial speculation, a fact that would seem to correspond to what the etymology already suggests, namely, that it is a matter of a specular game, a play of mirrors.[16]

Financial markets involve a radical form of uncertainty, an uncertainty that cannot be probabilized. Keynes shows that the only *rational* conduct in such a context is to imitate the others. Such a view clearly breaks with classical conceptions of the economic agent, for in economics and elsewhere rationality is generally associated with an autonomous form of decision-making, and not with being influenced by others. Yet this heterodox position is justified by Keynes for two reasons, one being quite general, the other being specific to financial markets. If I know nothing about the social context in which I find myself (as in the context of a panic), there is some chance that others may know something, and by imitating them, I may draw some advantage from their knowledge. Such is the general argument, but one hardly sees how it is supposed to apply to the case of an expert who trades on the financial markets. Yet Keynes asserts that this expert is in no better position than the "ignorant mass" of small fries trading on the market. Even if the expert notes a big difference between the "objective" or "fundamental" value that he attributes to a stock and its price on the market, he cannot afford not to take the latter into account, no matter how absurd it may seem. For if he finds himself obliged to *liquidate* his portfolio, he will have to do it at the market price, *volens nolens*. The only winner at this game is the person who can guess better than the crowd itself what the crowd is going to do.

Outguessing the crowd is the very essence of the contest in example seven, and this is why Keynes himself employed the same kind of example. The game mimics the breakdown of common sense, and the panic on the stock market is only a particularly striking example of the same thing. No longer able to base themselves on evaluations thought reliable because objective, that is, beyond both subjectivity and intersubjectivity, the agents are reduced to following each other's leads, everyone doing the same. Some, at least, are aware of it.

The decisive contribution made by Keynes resides in his understanding that the specular mechanisms that lead to crisis are the same mechanisms that permit its resolution.[17] In a case such as example

seven, if playing the game results in a solution, and if this solution "takes," as does a convention (in Lewis's sense) in a coordination game, then this solution is constitutive of a new order. In the financial context, this order would involve, for example, arriving at a new and stable set of financial evaluations. What is remarkable is that Keynes calls such solutions "conventions." He writes that "the psychology of a society of individuals each of whom is endeavoring to copy the others" leads to what can quite rightly be called a "conventional judgement."[18] But convention in Keynes's sense is not the same as what Lewis has in mind.[19]

Let us see how Keynes describes the *emergence* of a solution (whereas Lewis says nothing about the problem of morphogenesis). The basic schema is as follows. Suppose there are two subjects *A* and *B* who are engaged in reciprocal imitation. Now, let us imagine that a rumor makes *A* think that *B* desires (wants to purchase, longs for, etc.) an object, *O*. Henceforth, *A* knows what he must desire in turn (or wish to buy, or long for, etc.). Acting in keeping with this new-found desire, *A* designates to *B* that *O* is the object of his (*A*'s) desire. When *B* in turn shows interest in *O*, *A* sees this as proof that his initial hypothesis was right. Here we see the emergence of an objectivity or exteriority through the closure of a system in which all of the agents imitate each other reciprocally. This kind of objectivity can acquire an even greater strength in function of the number of agents involved. The most ridiculous rumors can lead a crowd to have a unanimous attitude towards the most unexpected object, for each individual in the group sees proof of the object's value in the actions and looks of the others. This kind of process has two moments. The first is the moment of the specular and speculative play of looks, when everyone scrutinizes the others for some sign of the coveted knowledge. Sooner or later, everyone is precipitated in the same direction. The second moment is that of the stabilization of the emergent object, a stabilization achieved through a forgetting of the arbitrariness inherent in the very conditions of that object's genesis. The unanimity that was responsible for its genesis projects the object outside the system of specularity for a certain period of time. The actors, who now all aim their glances in the direction indicated by the object, no longer stare at each other and their glances no longer cross.

Thus in Keynes, what stabilizes the solution or convention is the group's *misrecognition* (*méconnaissance*), and not, as in Lewis, CK. The contrast could not be more extreme. Here it is no longer possible to identify null and infinite specularities: the convention totally blocks the play of specularity. The figure of CK is at the horizon of the crisis and evaporates when the crisis is over.

3.3. The Constitution of the Public Space

The difference between a *public secret* and CK is best illustrated by H. C. Andersen's tale about the emperor's new clothes. Everyone sees that the emperor is naked but no one dares to imagine that the others see the same thing. This is why the emperor can go on being what he is. The truth is told by a young child, and it is the Other who is speaking through him. Suddenly, a *public space* is opened up and the political state of the society is radically changed, or better, it is created. This space is the sphere of promises, agreements, explicit arrangements, pacts, and social contracts.

Our fifth example illustrates the logic of the second state of nature that Rousseau evokes in his *Discours sur l'Origine et les fondements de l'inégalité parmi les hommes*. This is the bad natural society, the Hobbesian stage in Rousseau's reconstruction of the history of mankind. This is the stage into which the first state of nature collapsed, inevitably and irremediably, as the result of the slightest swerving or *clinamen*. The game in example five is opportune for giving a formal account of the argument in Rousseau's famous discussion of stag hunting.[20] Either men coordinate their efforts so as to hunt stags (C), or they hunt rabbits individually (B). Anyone who tries to hunt stags all alone will starve. As soon as specularity gets going, mankind enters the realm of *l'amour-propre*. Men live in the eyes of others. They become mutually enclosed within the bad natural equilibrium (BB). The transition to the good society of the contract (CC) remains an enigma in Rousseau's text, but the nature of the solution is fairly clear: artificial means must bring mankind back to the realm of *l'amour de soi* that characterized the first state of nature (let everyone vote on his own in a kind of absolutely private voting booth; and let everyone vote for himself while being guided entirely by the "preference that each

gives to himself").[21] Here we see the solution set forth in example five – null specularity.

Example six and Andersen's tale link the constitution of a public space to CK. An uttering of truth in public initiates transparency. Here we have a contractarian philosophy in John Rawls's sense, not Rousseau's. It is not surprising that the author of *A Theory of Justice* refers to Lewis and CK when he sets forth the condition of *publicity* that must be satisfied by his principles of justice. This condition plays a fundamental role in the demonstration that the contracting parties will, in the original position (Rawls's equivalent of the state of nature), prefer Rawls's principles of justice to the utilitarian principle.[22]

We began with Lacan and will end up with him as well. Vincent Descombes has quite rightly commented that there are two, not one, definitions of the symbolic in the structuralist's discourse.[23] According to the first definition, the symbolic is a "signifying convention," which means that it amounts to the conventional setting up of an arbitrary sign, this sign being the symbol of that convention (here we encounter the reflexivity of communicative intention, dear to Anglo-Saxon pragmatists). In this first sense, the symbolic is the mark and the effect of a collective will, a social contract. Yet the second definition leads to a paradoxical inversion: "When symbols are no longer the effects of convention, they are its source. They do not come after the social bond, they come before it and produce it." Now it is the symbol that creates mankind and society, and not the reverse.

The kind of specular game presented in example six had in Lacan's eyes the advantage of reconciling these two seemingly incompatible interpretations of the symbolic. The contractual dimension is present in the form of the utterance of the person who speaks in the name of the Other, thereby having the status of speaker of the truth. There is a pact of listening that no one can disobey because what the Other knows, everyone must know.[24] But these specular games were also supposed to illustrate, if not to ground, the thesis of the transcendent Other, the Other who governs the imaginary from without. Our own conclusion is rather different, and becomes clear when the game in example six is included alongside the full range of illustrations that we presented. Yes, the social pact has a subject, but far from being external or transcendent in relation to the members of society, this subject is itself the product of their pact.

NOTES

[1] One of the rare exceptions known to this author is Derrida's "deconstruction" of Austin, which lead to a polemical exchange with John Searle. See Jacques Derrida, 'Signature évènement contexte', in *Marges de la philosophie* (Paris: Minuit, 1972), 367–93; John Searle, 'Reiterating the Differences: A Reply to Derrida', *Glyph* **1** (1977), 198–208.

[2] I borrow this example and this analysis from Dan Sperber and Deirdre Wilson, *Relevance: Communication and Cognition* (Oxford: Basil Blackwell, 1986), 30–31, 61–62. These authors draw upon and criticize the ideas of Peter Strawson, 'Intention and Convention in Speech Acts', *Philosophical Review* **73** (1964), 439–60; and Steven R. Schiffer, *Meaning* (Oxford: Oxford University Press, 1972), Chapter 2. I must point out, however, that Sperber and Wilson do not use this example to lead to the idea of common knowledge, for they reject this notion as being "psychologically implausible." Rather, the hair dryer episode is used to illustrate the functioning of the alternative concept in their theory of communication, namely, that of "mutual manifestness."

[3] David Lewis, *Convention: A Philosophical Study* (Cambridge, Mass.: Harvard University Press, 1969).

[4] Pierre Bourdieu, *Esquisse d'une théorie de la pratique* (Geneva: Droz, 1972), 230. See also my analysis of this text in 'Totalisation et méconnaisance', in Paul Dumouchel, ed., *Violence et vérité: autour de René Girard* (Paris: Grasset, 1985), 110–35.

[5] I borrow this analysis from Serge-Christophe Kolm, *Philosophie de l'Economie* (Paris: Seuil, 1986), 305–06.

[6] R. J. Aumann, cited by Jean-Michel Lasry, 'Le Common Knowledge', *Ornicar* **30** (1984), 75–93.

[7] Lasry makes the interesting conjecture that the loop connecting the theoretician's theory to the theory of the agents could possibly lead to strange outcomes in the case of other games, including an absence or multiplicity of theories.

[8] In the game of the three prisoners as analyzed by Lacan in the previously cited article, the punctuation of time is no longer given to the players from without. In this case the standard solution is in danger of self-destructing. The rhythm has to be generated endogenously. Lacan's text is thus worthy of further examination.

[9] Paul Ricoeur, 'Aliénation', *Encyclopaedia Universalis* (Paris: Encyclopaedia Universalis, 1984), 769–73.

[10] This example is taken from J. Gould, cited by Lasry.

[11] Thomas C. Schelling, *The Strategy of Conflict* (Cambridge, Mass.: Harvard University Press, 1960).

[12] Cf. David Lewis, 'Languages and Language', in *Philosophical Papers*, Vol. 1. (Oxford: Oxford University Press, 1983), 166.

[13] The game is said to be "non-cooperative" because the agents do not communicate with each other. This is an unfortunate choice of expression, as Schelling notes, because what is at stake in the game is the cooperation (coordination) of the agents.

[14] See for example Tyler Burge, 'On Knowledge and Convention', *Philosophical Review* **84** (1975), 249–55.

[15] An example is David E. Cooper, 'Lewis on Our Knowledge of Conventions', *Mind* **86** (1977), 256–61.

[16] John Maynard Keynes, *Collected Writings*. Vol. VII: *The General Theory of Employment, Interest, and Money* (London: Macmillan, 1973), 156.

[17] Such is the reading set forth by André Orléan in his 'Mimétisme et anticipations

rationnelles: une perspective keynesienne', *Recherches Economiques de Louvain* **52**:1 (1986), 45–66.

[18] Keynes, 'The General Theory of Employment', *Quarterly Journal of Economics* **51** (1937), 209–223, esp. 217.

[19] Everything looks as if Schelling foresaw it (well before Lewis developed his theory). In a note, Schelling refers to the example of the beauty contest in Keynes and states that although it is exactly the same problem that he is dealing with, his conception of the solution is not at all the same as Keynes's (*Strategy of Conflict*, 94).

[20] Most frequently it is the prisoner's dilemma that is used for this purpose.

[21] See Lucien Scubla, 'Sur l'impossibilité de la volonté générale chez Rousseau', *Cahiers du CREA* **1** (1982), 69–137.

[22] John Rawls, *A Theory of Justice* (Oxford: Oxford University Press, 1971), 133.

[23] Vincent Descombes, 'L'Equivoque du symbolique', *MLN* **94** (1979), 655–75.

[24] Lasry, 'Le Common Knowledge'.

CREA Ecole Polytechnique,
1, Rue Descartes,
75005 Paris,
France.