Common Knowledge and Interactive Behaviors: A Survey [1]

Frédéric KOESSLER

BETA – Theme
Université Louis Pasteur, 61 Avenue de la Forêt-Noire
F-67085 Strasbourg Cedex (France)
E-mail: `koessler@cournot.u-strasbg.fr`
Tel: (33) 3 90 41 40 69 – Fax: (33) 3 90 41 40 50

May 26, 2000

Abstract

This paper surveys the notion of common knowledge taken from game theory and computer science. It studies and illustrates more generally the effects of interactive knowledge in economic and social problems. First of all, common knowledge is shown to be a central concept and often a necessary condition for coordination, equilibrium achievement, agreement, and consensus. We present how common knowledge can be practically generated, for example, by particular advertisements or leadership. Secondly, we prove that common knowledge can be harmful, essentially in various cooperation and negotiation problems, and more generally when there are conflicts of interest. Finally, in some asymmetric relationships, common knowledge is shown to be preferable for some players, but not for all. The ambiguous welfare effects of higher-order knowledge on interactive behaviors leads us to analyze the role of decentralized communication in order to deal with dynamic or endogenous information structures.

KEYWORDS: Interactive knowledge, common knowledge, information structure, communication.

JEL CLASSIFICATION: C72, D8.

Résumé

Cet article présente la littérature sur la notion de connaissance commune et met en évidence le rôle des connaissances interactives dans les problèmes économiques et sociaux. La notion de connaissance commune utilisée et analysée est celle provenant de la théorie des jeux et de l'informatique théorique. Nous montrons tout d'abord qu'une base de connaissance commune est souvent utile, voire nécessaire à la coordination, aux accords et aux consensus entre agents, et à l'obtention d'un équilibre. Différents mécanismes et contextes permettant de générer une connaissance commune sont présentés, dont la publicité et le leadership. Nous montrons ensuite que la connaissance commune peut s'avérer néfaste des contextes stratégiques comme ceux de coopération ou de négociation, et plus généralement lorsque des conflits d'intérêt apparaissent. Enfin, dans des relations asymétriques, la connaissance commune peut avoir des effets bénéfiques pour certains agents, mais négatifs pour d'autres. Les effets souvent ambigus des connaissances interactives sur le bien être nous mènent finalement à considérer une communication décentralisée et des structures d'informations dynamiques, voire endogènes.

MOTS CLÉS : Connaissances interactives, connaissance commune, structure d'information, communication.

CLASSIFICATION JEL : C72, D8.

Introduction

When an agent communicates some knowledge to an other, the latter integrates this new knowledge and combines it with his own. Moreover, new knowledge from the external world may be acquired when an agent observes external phenomena and events, experimental and statistical data, expert testimony, scientific tests or publicity campaigns. But that is not all an agent can learn. The aim of this paper is to show that in interactive situations, not only such individual and independent knowledge of "fundamentals" (first-order knowledge) is important, but also *interactive knowledge* (higher-order knowledge), i.e., knowledge about others' knowledge. If communication were simply a knowledge transfer from some agents to other agents, the first ones would never change their behavior when they reveal their knowledge, because they act according to their own knowledge which would not be modified. In the same way, an agent would be indifferent to reveal or not his knowledge because this act might not change his own knowledge. Such a point of view is obviously false: our behaviors depend drastically on our knowledge about others' knowledge. Others' knowledge being part of knowledge itself, we also act according to the knowledge we have about others' knowledge about our knowledge, and so on. When such a reasoning is applied infinitely and leads to the same knowledge for every agent, this kind of knowledge is called *common knowledge*.

To put it more precisely, a knowledge is common among a group of agents if everyone has it, everyone knows that everyone has it, everyone knows that everyone knows that everyone has it, and so on *ad infinitum*. Actually, it is simply an infinite regress of reasoning about agents' knowledge which unifies the set of distributed knowledge. It makes collective knowledge perfectly transparent to each individual. This definition of common knowledge requires infinite-order knowledge when the common parlance definition only requires first-order knowledge (i.e., knowledge about external states without interactive knowledge). This last knowledge is called *mutual knowledge*, and applies when a fact is known by every agent.

The notion of common knowledge was first proposed by Lewis (1969) in the philosophy of language, in a context of conventions, which could be seen as some common knowledge tools useful for coordination.[1] That is, coordination requires infinitely many levels of knowledge. Lewis attributes this idea to Schelling (1960), even if Littelwood (1953) already gave some examples using common knowledge reasoning. Common knowledge was formalized for the first time in economics by Aumann (1976), based on the developments about states of the world representations by Savage (1954) and Harsanyi (1967–1968), but without reference to strategic or market behaviors.

In this paper, we review classes of well known interactive problems where agents' behaviors are driven by the configuration of interactive knowledge and by common knowledge assumptions. More precisely, this paper deals specifically with the differences between mutual, distributed, individual, common, and "almost" common knowledge, how these levels could be generated, and how they influence interactive behaviors. Making a rigorous distinction between different levels (or hierarchies) of knowledge is crucial to study problems such as coordination, cooperation, negotiation, trade or betting, and to understand the impact of communication and knowledge sharing.

The paper is organized as follows. In section 1, we discuss and describe the concepts of knowledge and common knowledge taken from game theory (and economics) and computer science. In the sections that follow, we consider interactive decision making, and we study the impact of common and interactive knowledge on strategic and decentralized

---

[1] Think about the road code for motorists. For more discussions about the links between conventions and common knowledge see, e.g., Dupuy (1989).

1

behaviors. We first show in section 2 that common knowledge is essential to solve coordination problems. The example of the electronic mail game will easily show the relevance of common knowledge, and that it is far from being an innocuous proxy for shared or mutual knowledge. Afterwards, in section 3, we show that common knowledge can actually have negative effects in presence of conflicts of interest. In asymmetric relationships, common knowledge can also be preferable for some players, but not for all. In particular, we show in section 4 that often potential "winners" effectively "win" for sure only if the interactive situation is common knowledge. All this analysis of negative, positive and asymmetric effects of common knowledge is not based on pure, first-order information effects, where we know that almost all situations are conceivable,[2] but on interactive knowledge effects, with fixed first-order knowledge. For example, when we consider rationality, our interest is not on the effects of departures from rationality or from mutual knowledge of rationality, but only of a lack of common knowledge of rationality. Finally, in section 5, we show how common knowledge can evolve through decentralized interactions. Because of ambiguous welfare effects of interactive knowledge, we especially focus on the role of *decentralized* communication on learning and equilibrium outcomes to get dynamic or endogenous information structures. Two kinds of relevant configurations drive the evolution of the information structure. One is linked to certification possibilities and language richness. The other is related to communication practices and strategic contexts in which agents interact.

1   The game-theoretic and computer science approaches

In this section, we present the standard modeling of knowledge used in game theory, and we relate it with the computer science approaches. To begin with, the basic conceptualization of knowledge and belief is exposed. Afterwards, we detail the basic framework used to represent an agent's knowledge. The standard axioms of knowledge are discussed. Then, we extend the analysis to many agents in order to take into account interactive knowledge. Finally, we show how the configuration of knowledge can be represented in terms of information structures, as it is usually the case in economic theory.

1.1   Overview

In short, the game-theoretic approach of knowledge is especially based on probabilities and sets theories, whereas the computer science approach is based on the mathematics of logic, especially non-classical applied logic such as epistemic or modal logic.

In game theory, information is represented by an *event*, i.e., a set of *states of the world*. A state is a 'complete description of the world, leaving no relevant aspects undescribed' (Savage (1954, p. 9)). An agent knows an event if he believes it with probability one.[3] In this case, the event must be realized (the agent's knowledge is true). An agent believes an event at 90% if he assigns probability 0.9 to this event. Then, this event will occur with probability 0.9 *in his mind* (other agents may have different beliefs). The evolution of beliefs and knowledge follows the *Bayesian learning rule*, based on conditionalization. There is no real cognitive difficulty in learning, because isolated agents cannot learn by thinking. They immediately learn if they receive messages, if they observe external phenomena, or if they observe others' behaviors. In particular, a set of agents cannot learn more than what is distributed among them. Therefore, learning is "passive" in the sense that there

---

[2]Simple examples are given in Bassan, Scarsini, and Zamir (1997).

[3]We suppose that the set of states of nature is finite. If not, the technical analysis becomes more complicated, without many conceptual differences.

is no real discovery about the environment. This does not mean that new knowledge, not known before, cannot be learned. Agents may not only sum their knowledge, but they can also combine it to discover new realized events. It this way, they could also discover the knowledge of the others, which can lead to common knowledge of some facts (more about such an evolution is given in section 5). In this sense, knowledge could in no manner be assimilated as a good, a capital, a cost, or a real capacity. Knowledge is formed by agents' anticipations, beliefs, information and representations.

The computer science representation of knowledge takes its grounds from the first epistemological treatment of knowledge in philosophy, during the sixties. Hintikka (1969) was the first to introduce the notions of knowledge and belief operators. The abstract model theory is often based on *epistemic logic*, using logical symbolic operators (such as the knowledge operator), also called proportional connectors, associated with truth values. The most common semantics associated to such language are Kripke's (1963) semantics, based on *possible worlds*. In such an approach, an agent knows that a proposition is true iff it is true in every world the agent conceives as possible. As in game theory, a lot of applications in cognitive sciences, linguistics, artificial intelligence and psychology permit to extend the analysis to more sophisticated reasoning with many interacting agents, and with a not necessarily static environment.[4]

These approaches have much in common. With the usual assumptions made in game theory, and barring the probabilistic features, they are even equivalent. Generally, the computer science approach is less restrictive on knowledge and reasoning, and does not always use probabilistic knowledge. But game theory has the advantage to deal with strategic behaviors. In fact, in game theory agents are strategic players, although in computer science agents represent generally physics processors or robots.

1.2 A formal knowledge representation

In this subsection we present the more usual knowledge modeling of game theory, and we relate it to computer science modeling. Such a modeling is based on a set of states of the world, and agents' information is characterized by partitions (i.e., sets of mutually exclusive subsets including the whole set of states) of the state space. There is no unique and "correct" modeling. The following, which is sometimes called *view-based* or *state-based*, is very tractable and has the advantage to be applicable to the majority of economic problems dealing with uncertainty. More important, and for the purpose of this paper, it allows us to take into account interactive knowledge, and thus common knowledge.

Let $\Omega = \{\omega_1, \ldots, \omega_m\}$ be the finite set of possible states of the world, or states of nature, also called fundamental space. This set represents all non-controllable uncertainty. The probability distribution on Ω is denoted by p, and $N = \{1, \ldots, n\}$ is the (finite) set of agents.

A subset $E \subseteq \Omega$ is an event. We say that this event is realized if the true state ω is in E. To each event we can associate a literal meaning, i.e., each event could have a semantic content. For example, if $\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4, \omega_5, \omega_6\}$ represents the possible results of a dice cast (where ω_i means that number i has been drawn), the event $\{\omega_1, \omega_3, \omega_5\}$ could mean "the realized number is odd". The event Ω, including all possible states of nature, is necessarily always true.

In computer science models devoted to the analysis of distributed systems, the propositions, or formulas, are expressed in a formal language. They belong to a set of propositions closed by negation, conjunction, and by the application of knowledge operators.

---

[4]See, e.g., Halpern (1995) or Fagin, Halpern, Moses, and Vardi (1995) for some presentations.

Each proposition is obtained from primitive propositions by a combination of the following transformations: negation, conjunction, disjunction, and implications of formulas. These last ones respectively correspond to complement, intersection, union, and inclusion of events. A binary function could associate to each state and each formula an element of the set $\{True, False\}$. Such a knowledge structure, represented by a propositional model, is called a *Kripke structure*. It is a semantic approach, whereas the approach with a formal language without states of nature is a syntactic one. For more details, see Aumann (1999a, 1999b) who relates these two approaches. He shows how to construct a semantic model, starting with a syntactic one, by constructing a canonical system which associates states of nature to propositions. This allows him to show explicitly that the semantical knowledge model is itself common knowledge. Walliser (1991) describes the two approaches by making links with game theory. What is interesting with such modeling is that it highlights the *logical omniscience* problem: agents, knowing all the tautologies[5] (i.e., all the basic, logical, and necessary true and satisfied formulas), also know by direct deductions, substitutions and replacements, all theorems and logical consequences of their knowledge.

The logical omniscience problem is actually linked to the fact that we work with a closed system. This is also found in *information theory*, which has in fact much in common with game theory, because applied in many fields and because of both a mathematical and a conceptual component. Indeed, Shannon's (1948) theory of information applies to finite and probabilistic worlds.[6] Thus, knowledge evolution is characterized by a *diminution* of uncertainty. It could be justified in the short term and in local analysis, but in the long term it is clear that knowledge exploration and evolution is done through an *expansion* of uncertainty. The system of knowledge is closed and fixed. That is why such theories should not be applied to too general problems.

Another common limitation of game and information theories is that some probabilities come before information, which is sometimes difficult to justify: to know that the probability to get a certain face in a dice cast is 1/6, you have in fact to receive some information to identify a face appearance. The advantage of game theory is however that information has a semantic content: a message informing you if you buy a good quality or a bad quality product could contain the same *quantity* of information, but will certainly generates markedly different behaviors. Another advantage of game theory is that it takes into account metainformation (i.e., higher-order information), contrary to Shannon's information theory.

1.3 An axiomatics of the knowledge operator

A useful tool to analyze interactive knowledge (or recursive belief) of many agents is to describe separately their knowledge by a simple operator. We denote by K_i the *individual knowledge operator* of agent i. In this model, it is a function,

$$K_i : 2^\Omega \to 2^\Omega,$$

where the power set 2^Ω is the set of subsets of Ω. The set K_iE is the event "agent i knows E".

To be tractable, the knowledge operator often has to verify four usual (albeit questionable) axioms that we will comment one after one. Many implicit and hidden assumptions

---

[5]This name comes from the philosophy literature.

[6]Game theory applies to infinite and even uncountable states space, but a measurable space is necessary, i.e., one must define a σ-field on the state space and a probability measure on it.

in economics and game theory are behind these axioms. The event $\bar{E}$ will denote the complement of E in Ω. Let $\bar{K}_i$ be the operator of ignorance such that $\bar{K}_i E$ is the complement of $K_i E$ (i.e., $\bar{K}_i E$ means "agent i does not know E").

(A1) $K_i(E \cap F) = K_i E \cap K_i F$ (axiom of deductive closure, or *distribution axiom*): an agent knows E and F iff he knows E and he knows F. One can easily see that with this axiom, if $E \subseteq F$ then $K_i E \subseteq K_i F$, i.e., if E implies F and the agent knows E, he also knows F. This property characterizes the logical omniscience phenomena.[7]

(A2) $K_i E \subseteq E$ (non delusion, knowledge, or *truth axiom*): all what an agent knows is true. This axiom allows a theoretical (but objectionable) distinction between knowledge and belief.[8] It can be decomposed into three properties. If E is an event which only characterize "fundamentals" (and not knowledge about fundamentals), this axiom is equivalent to $K_i K_i E \subseteq K_i E$ (true knowledge about personal knowledge), $K_i \subseteq E$ (true knowledge about "fundamentals") and $K_i K_j E \subseteq K_j E$, $i \neq j$ (true knowledge about others' knowledge). These properties are clearly ordered from the most to the least justifiable ones.

(A3) $K_i E \subseteq K_i^2 E$ (transparency, or *positive introspection axiom*): if an agent knows E, then he knows that he knows E. The reciprocal is obtained by (A2), so (A2) and (A3) give $K_i E = K_i^2 E$. By this axiom, agents are always aware of what they know.[9]

(A4) $\bar{K}_i E \subseteq K_i \bar{K}_i E$ (axiom of wisdom, or *negative introspection axiom*): if an agent does not know E, then he knows that he does not know it. It is the more restrictive axiom because it does not allow agents to ignore their ignorance (they are conscious of their ignorance).

The last two axioms give knowledge an internal coherence and preclude unawareness:[10] agents can inspect their own base of knowledge and distinguish what they know from what they do not know. This is acceptable if we suppose that, before making their decision, they seek to recognize and to distinguish their ignorance. The negative introspection axiom is the most demanding one. Shin and Williamson (1994) even show that the knowledge of Turing machines is not consistent with it, but can satisfy all the other axioms. By the preceding axioms we get the following property:

(A0) $K_i \Omega = \Omega$: an agent always knows that the universal event Ω is realized. In other words, he knows that all the possible states of nature belong to this event. Indeed: $K_i \emptyset = \emptyset$ by (A2), and $K_i \Omega = K_i \bar{\emptyset} = K_i \bar{K}_i \emptyset = \bar{K}_i \emptyset = \bar{\emptyset} = \Omega$ by (A4) and (A2). Individuals can thus never be surprised by an unforeseen event; they always know in advance what they can possibly learn.

To accept this last property it is necessary to imagine a relatively stable context, with which agents are familiar, which enables them to solve complex problems. Perfect rationality is an acceptable approximation when agents are experimented. Think for example

---

[7] For alternative approaches and some discussions see, e.g., Gillet and Gochet (1993) and Fagin, Halpern, Moses, and Vardi (1995, chap. 9).

[8] When Ω is finite, belief with probability one is equivalent to knowledge. Brandenburger and Dekel (1987) generalized Aumann's (1976) work with infinite state space with null event (i.e., events with measure zero).

[9] To justify this axiom, think about a chess game: if we would not know what we know, we would be able to play chess against ourselves, but this has in fact no sense.

[10] See, e.g., Modica and Rustichini (1994, 1999) for the introduction of unawareness by assuming away the negative introspection axiom.

of an internal firm's organization where individuals are familiar with the risks they incur, or of an insurance or a bank which faces relatively known risks.

A knowledge system verifying properties (A0) to (A4) has been called *S5 system* by Kripke (1963) and *strong epistemic model* by Bacharach (1985).

1.4   From individual knowledge to higher-order knowledge

To illustrate the force of the previous axioms, we can show that individuals' knowledge is necessarily very broad, even for those who apparently do not have much of it. Indeed, suppose (see Fagin, Halpern, and Vardi (1991)) that an agent A, Alice, is only informed about the event E. Logically, she also knows all its consequences, but it is not all. Let F be another event not having any apparent link with E. Alice does not know this event. But according to the axioms of introspection, she knows she does not know F: all that she is supposed to know is E, and yet she also knows $\bar{K}_A F$, which is not a logical consequence of E! Worse, suppose that Bob, the agent B, joins Alice. Then, we can notice that Alice knows that Bob does not know that Alice knows F (she knows that he does not have false knowledge). In the same way, Alice knows that Bob knows this fact (that he does not know that Alice knows F) since she knows that Bob can also make introspections. Accordingly, we have $K_A K_B \bar{K}_B K_A F$, and thus, in spite of her "weak knowledge", she knows a non-trivial fact about Bob's knowledge! Consequently, for these axioms to be reasonably accepted, it is necessary that Ω describes a "small world",[11] i.e., that the set of possibilities is sufficiently small, so that potential surprises could be discounted in advance, and each one could evaluate all implications of what can happen. The set Ω must be stable, all possible states of nature must have a positive probability at the beginning. The advantage in the description of a game is that this assumption is verified, since the rules of the game describe this "small world".

The knowledge operator and its property being defined, we now turn to the description of interactive and common knowledge. Readers interested in the knowledge modeling presented here are referred to Brandenburger and Dekel (1989) and Binmore and Brandenburger (1990) for non-technical surveys, and to Geanakoplos (1994) and Battigalli and Bonanno (1999) for more technical ones.[12] The latter makes particularly links between this modeling, modal logic, and various solution concepts in game theory.

Let K be the *mutual knowledge operator*, such that KE is the event "everybody knows E". We will see that mutual knowledge does not have the same nature as common knowledge. The event KE can naturally be expressed using the individual knowledge operators as follows:

$$KE = \bigcap_{i=1}^{n} K_i E.$$

Similarly, the event "everybody knows that everybody knows E" is given by,[13]

$$K^2 E = K \bigcap_{i=1}^{n} K_i E.$$

---

[11]This assumption was introduced by Savage (1954), and was taken up again by Binmore (1992, p. 448).

[12]See also the special issues of *Theory and Decision 37(1), 1994* and *International Journal of Game Theory 28(3), 1999*, and the book of Bacharach, Gérard-Varet, Mongin, and Shin (1997).

[13]Such second-order mutual knowledge is similar to the notion of "collective knowledge" (Marengo (1992), Crémer (1993)). This literature on organizational learning does not focus on interactive knowledge. Nevertheless, it is implicitly present when Crémer (1993) gives the example of the janitor in a firm: the place where cleaning material is stored is only known by the janitor, but the fact that he has this knowledge is part of corporate culture (collective knowledge). Knowing what the others know includes at least one level of interactive knowledge.

More generally, the event "everybody knows that [k times] the event E is realized" is defined by $K^k E$. When $\omega \in K^k E$, the event E is also said to be *known up to a level k* at ω. If $k \geq m = \#\Omega$ we can show that we have

$$K^k E = K^m E. \tag{1}$$

Indeed, by (A2) we have $K^m \subseteq K^{m-1} \subseteq \cdots \subseteq KE \subseteq E$; if there exist $r \in \{0, \ldots, m\}$ such that $K^r E = K^{r+1} E$ then this equality is also verified for $k \geq r$ and in particular for $k \geq m$. On the contrary, $\#K^m E < \#K^{m-1} E < \cdots < \#KE < \#E \leq m$. Thus, $\#K^m E = 0$, i.e., $K^m E = \emptyset = K^k E$ for $k \geq m$.

Then, the event "E is common knowledge" could be defined by

$$CKE = \bigcap_{k=1}^{+\infty} K^k E.$$

By equation (1) and axiom (A2) we get

$$CKE = \lim_{k \to +\infty} K^k E = K^m E.$$

Note that when Ω is infinite and countable, as it will be the case in the "electronic mail game", then there is not necessarily a finite number m such that $CKE = K^m E$. In this case, when m is very large but finite, the event E will be called "almost" common knowledge if $\omega \in K^m E$ but $\omega \notin CKE$.

We can easily prove that common knowledge, given the axiomatics of subsection 1.3, verifies the following properties:

- $E \subseteq F \Rightarrow CKE \subseteq CKF$: if an event induces another one, common knowledge of the first one also induces common knowledge of the second one.

- $CKE = K_i CKE, \ \forall \, i \in N$: an event is common knowledge for a group of agents iff each agent of this group knows that it is common knowledge.

- $CKE = CK(CKE)$: an event is common knowledge iff it is common knowledge that this event is common knowledge.

- $CKE \subseteq E$: if an event is common knowledge, it is true. More generally, higher-order knowledge always implies some lower-order knowledge.

- $CK\Omega = \Omega$: the event "one of the states of nature is realized" is common knowledge.

1.5 Knowledge and information structures

Knowledge verifying the preceding properties has the convenience that it can be defined in terms of *information partitions*. Proofs are given, e.g., by Binmore (1992) or Geanakoplos (1994).

Intuitively, an agent knows E if this event is verified at all the states of the world he considers possible. The set of states that an agent i considers possible when the true state is ω (in other words, the set of states where he does not know that ω does not realize) is given by

$$h_i(\omega) = \bar{K}_i \bar{\omega}.$$

This set is called the *information set* of agent i at ω. Intuitively, it characterizes i's state of mind at ω. We can show that $(h_i(\omega))_{\omega \in \Omega}$ constitutes a partition of Ω for all

$i \in N$ because $h_i(\omega)$ and $h_i(\omega')$ are either disjoint or identical and because $\omega \in h_i(\omega)$. The individual knowledge operator can be deduced from information sets by the following unique relation, which means that agent i knows E iff all states he considers possible are included in E:

$$K_i E = \{\omega \in \Omega : h_i(\omega) \subseteq E\}. \tag{2}$$

Thus all theoretical analyses based on information partitions implicitly assume axioms (A1) to (A4). Usually, information structures used in economics are even less general because they do not include interactive knowledge: when some agents know something, it is common knowledge that they know. Often models either assume that some information is strictly private, or they assume that it is common knowledge, without considering heterogeneous intermediate interactive knowledge. In other words, events in the form $K_i K_j \cdots K_k E$ do not depend on the agents $i, j, \ldots \in N$, i.e., everybody has the same knowledge about any agent k's knowledge.[14] To question this axiomatics leads then to criticize a large part of development in game theory, decision theory, and consequently in standard economics.

By equation (2), the event "everybody knows E" is

$$KE = \{\omega \in \Omega : \bigcup_{i=1}^{n} h_i(\omega) \subseteq E\},$$

and the event "E is common knowledge" is

$$CKE = \{\omega \in \Omega : \bigcup_{i=1}^{n} h_i(\omega) \subseteq K^{m-1}E\}.^{[15]}$$

Without going into technical details, it is important to point out that one can formulate the definition of common knowledge in a non-recursive manner, by using the notion of the *Meet* of partitions. It is the finest common coarsening of all partitions, and is written $\mathbb{M} = H_1 \wedge \cdots \wedge H_n$. By denoting $M(\omega)$ the element of $\mathbb{M}$ containing ω, one can prove that $CKE = \{\omega \in \Omega : M(\omega) \subseteq E\}.^{[16]}$ The set $M(\omega)$ contains all commonly possible states at ω. Such a definition of common knowledge does not use infinite regress of knowledge reasoning and it explains how common knowledge could be achieved without agents using an infinite regress of thought. Common knowledge in terms of public event and other definitions are given by Barwise (1988) who compares and gives equivalence conditions for these approaches. As we will see in the next section, common knowledge can arise all at once by public announcement.

Another extreme configuration of knowledge is *distributed knowledge*. We say that the knowledge about an event is distributed among the agents if it is known by the group when all agents combine and share all their knowledge. In other words, distributed knowledge is the knowledge *implicit in the group*. Formally, the event "the knowledge about E is distributed" is the event

$$DE \equiv \{\omega \in \Omega : J(\omega) \subseteq E\},$$

where $J(\omega) \equiv \bigcap_{i \in N} h_i(\omega)$ is the element of the *Join* (coarsest common refinement) of the agents' partitions containing ω. Therefore, distributed knowledge is the "dual" of common knowledge. We always have the following inclusions:

$$CKE \subseteq KE \subseteq K_i E \subseteq DE, \ \forall \ i \in N.$$

---

[14]See Morris, Postlewaite, and Shin (1995) for a formal characterization of the depth of knowledge in an information structure.

[15]m is the cardinal of Ω.

[16]See Aumann (1976) and Nishihara (1991). For an axiomatics directly defined on the common knowledge operator see Milgrom (1981a).

Therefore, a common knowledge is a mutual knowledge, which is an individual knowledge for each agent, which is a distributed knowledge. With heterogeneous agents the inverse inclusions are not verified.

2 Positive effects of common knowledge and coordination problems

Having a clear definition of common knowledge and of its properties, we now show how it is essential to solve, or at least to help to solve, coordination problems. Coordination is a very extensive and fundamental phenomenon in social interactions, and coordination problems are present even with conflicts of interest, because they often come prior to various conflicts. In this section we begin to illustrate the relevance of common knowledge in the electronic mail game, which is a typical coordination problem. The absence of strict common knowledge will forbid coordination, even if we have "almost" common knowledge. Afterwards we detail some mechanisms that create common knowledge and help coordination, particularly leadership and advertisement. Finally, in the last subsection, we show how common knowledge and coordination conditions can be weakened in order to modify the sometimes paradoxical electronic mail game solution. Many historical and empirical situations where common knowledge is important are given in Chwe (1998b, 1999a).

2.1 The electronic mail game

An important aspect in *face-to-face communication* is that when we teach, we also learn, because we acquire higher-order information. Face-to-face communication can even create infinite hierarchies of knowledge. To prove the implications of such metaknowledge in behaviors, one has just to consider the case of non face-to-face communication with *noisy communication*.

The *electronic mail game* (or *coordinated attack problem*) will easily illustrate that point because common knowledge will drastically change players' behaviors, compared with the case where there is no common knowledge, but only "almost" common knowledge (i.e., a large but finite order of mutual knowledge).

This game was introduced by Gray (1978), and analyzed in more details by Rubinstein (1989), Halpern and Moses (1990), Fagin, Halpern, Moses, and Vardi (1996), Morris and Shin (1997) and many others. In its origin, it was described as follows. Two generals, each commanding a division of an army, want to attack a common enemy which is posted in a valley between them. Only one general knows if the enemy is prepared to fight or not. They will win the battle only if they both attack simultaneously and if the enemy is unprepared. To coordinate an attack, the first general could send a messenger (or an e-mail in the "modern" version of the problem) to the other general, informing him whether the enemy is prepared or not. Unfortunately this communication method is not perfect, and nobody is completely sure that the message reaches the receiver.

We will naturally say that the coordination problem is solved if the three following conditions are verified: the attack never happens when the enemy is prepared, the two generals never attack alone, and the two generals both attack simultaneously if the enemy is not prepared. Such a perfect coordination proves never to be possible with an imperfect communication system, for every possible action rule (even not strategic). This is because imperfect communication cannot make common knowledge the fact that the enemy is unprepared.[17] Intuitively the reasoning could be described as follows. Suppose that the

---

[17]See Halpern and Moses (1990) in a computer science setting, and Rubinstein (1989), Heifetz (1996),

9

first general is informed that the enemy is unprepared, and then that they could coordinate a join attack. He sends a message to the other general. Suppose it reaches him, as it will almost all the time. The sender does not know if the message has arrived, so he cannot attack because he is not sure that the other will. Therefore, the second general sends back the message to confirm that he got the first one. Could they now attack, i.e., is it now sufficient that the second general receives the message and that the first knows that he receives it? The answer is "no" because now the second general is not sure that the first gets confirmation and then that he will attack, because he does not know if the first knows that he received it. We can continue this reasoning for any number of messages sent back and forth and see that they will never attack: the message must be common knowledge between the generals. What is important is not that the message is received, but that both know it is received, both know that both know it is received . . .

The same reasoning applies for more general coordination problems with many agents. The problem of rebelling against a regime is an example. Each agent rebels only if the others also do so. That is why communication for such collective actions takes the form of mass meetings, publications, flags and graffiti so that each one is aware that each one gets the messages, each one is aware that each one is aware that each one gets the messages . . . Every level of metaknowledge is called upon. Another coordination problem arises when people have to choose social goods (i.e., goods which are all the more attractive they are used more) or between different standards. In this case, advertisement will be the way to create the common knowledge condition, as we will see in subsection 2.3. Coordination problems are actually present in many fields, social movements, technological change, macroeconomics politics, and various organizations. The effects of higher-order knowledge on coordination in financial markets was also recently explored (see, e.g., Allen and Morris (1999)). As in the "electronic mail game", coordination in decentralized markets is very sensitive to a lack of common knowledge, and sufficient common knowledge is required to support coordination in risky outcomes.

Perfect coordination is however too demanding, and sometimes we only need that a good coordination happens almost all the time. In the last subsection we will see how to resolve this paradox. The problem seems unrealistic, and in practice coordination will be possible even with noisy communication, i.e., without common knowledge.

2.2   Leadership

We have seen that common knowledge is sometimes necessary for coordination but that decentralized noisy communication may not allow its creation. A simple and efficient manner to create common knowledge and to solve coordination problems is *leadership*. It was proposed by Foss (1999), who defines it as the ability to resolve social dilemmas by influencing beliefs. He uses a lot of ideas taken from game theory, also developed here. The originality of this work relies on the fact that he does not think that the only aim of the leader is to resolve incentive conflicts (for example, making agents cooperate in a prisoner's dilemma),[18] as it is usual in the economics literature of organizations.

The preceding subsection could be an explanation of why communication in organizations does not happen too much by telephone, fax, or e-mail, which are pairwise and non-public communications, often subjected to failures. Indeed, common knowledge could not evolve through noisy communication (as it was shown before) and evolves slowly in network communication (as we will show in subsection 5.1). Here, we show that a leader, who could reunite any kind of group in an organization, may influence members' beliefs simulta-

---

and Koessler (2000) in a game-theoretic setting.

[18]See section 3 for more about cooperation games.

neously and publicly through, for example, face-to-face contacts in meetings. By so doing, he could easily, cheaply and quickly (compared with convention formation, decentralized communication, adaptation and evolution) create some desired common knowledge. It is to be noted that the leader is not formally modeled. This theory relies on explanations similar to the focal point concept of Schelling (1960).

First, to achieve a specific equilibrium, coordination is actually necessary in games with incomplete information. If we consider for example the electronic mail game, leadership may allow a move from an inferior equilibrium (the equilibrium without common knowledge, where the generals do not attack) to the equilibrium where they coordinate an attack, by transforming the game into a *complete information* game (where it is common knowledge that the enemy is unprepared).

Secondly, leadership could be a method of *equilibrium selection*. Consider for example the game of figure 1. There are three pure Nash equilibria: (T,T), (M,M), and (B,B). The game is actually without incomplete information, so it is common knowledge, contrary to the electronic mail game. The leader could call players to play (T,T) or (M,M), so it becomes common knowledge that one of these equilibria is to be played. It this case, the leader does not make external events common knowledge (as it could be the case in the electronic mail game) but he makes each others' best response strategies common knowledge.

	Player 2		
	T	M	B
T	$(3,3)$	$(0,0)$	$(2,2)$
M	$(0,0)$	$(3,3)$	$(2,2)$
B	$(2,2)$	$(2,2)$	$(5/2,5/2)$

(Player 1 labels rows T, M, B.)

Figure 1: A coordination game with two symmetric Pareto dominant equilibria.

One could think that for such a game a leader is not necessary because (T,T) and (M,M) are strictly preferred to (B,B) by both players, and then (T,T) or (M,M) would be naturally selected without the leader's intervention. But this is not the case: game theory cannot evidently select a unique equilibrium with a general and valid solution concept based on individual rationality, even with cheap talk, and thus remains with a kind of "invisible hand" explanation.[19] One could always explain some intuitive selection with the notion of "focal point".[20] The so-called Stackelberg heuristics is another solution but, once again, it is not based on individual rationality but is a psychological explanation. Thus, leadership is a method of equilibrium selection because one has to choose between the prudent strategy yielding to the payoffs $(5/2,5/2)$ and the Pareto dominant strategies leading to the payoffs $(3,3)$.

Finally, if these last Pareto dominant strategies are selected, one has to coordinate on (T,T) or on (M,M). Therefore, a third aim of the leader is to allow players to *reach equilibrium* simply and cheaply. Indeed, even if game theory often supposes that behaviors are at equilibrium, this is not always practically true. The two Nash equilibria (T,T) and

---

[19]See van Damme (1995) who showed this fact even for symmetric team games, which are particular coordination games where players' payoffs are the same. In particular, payoff dominance is not justified from standard game-theoretic rationality assumptions. For cheap talk see, e.g., Farrell (1988) and Rabin (1994).

[20]When there is no obvious "focal" point, i.e., equilibrium strategies profile whose labeling has cultural significance or correspond to salient options, Schelling (1960) proposed that an external agent creates one by making a suggestion, and thus an agreement. As a matter of fact, what makes this external actor is what we call leadership.

(M, M) are perfectly symmetric, so it might happen that player 1 plays T and the other M. Making the equilibrium common knowledge, the leader helps to achieve equilibrium and, it this case, may be seen as a "Nashian regulator".

Leadership could be particularly efficient because players could trust the leader, and even follow its order, *if it is in their interests* (this is important because we are not allowed to control directly players' behaviors). The leader could be viewed as a *mediator*. By the way, such idea is already present in the concept of correlated and communication equilibrium,[21] but in a more restrained view, because the leader may have some reasoning so that he can make suggestions on his own initiative. Furthermore, because communication channels are integrated in the concept of correlated equilibrium, the mediator generates endogenous uncertainties not considered here.

Practically, the leader could be viewed as a top management who organizes meetings, rituals, ceremonies or media events where all the members are present simultaneously. These practices could also be viewed as *cultural practices*. To understand them, Chwe (1999a) also focus on common knowledge in the game theory sense and on coordination problems defined as situations in which a group of people would all gain by participating in some collective action, but in which no individual wants to participate unless other individuals in the group also participate. His argument is almost the same as those given before: face-to-face and public communication takes place when coordination problems must be solved, because it creates common knowledge in a simple way. Face-to-face conversation or eyes contact is important because if the receiver or the sender is away from his partner one could not verify if he pays attention, understand, is awake, knows that he knows that he understands ...[22,23] With such practices, knowledge is more transparent than without face-to-face communication. When there are more than two agents, but not too many, such face-to-face interaction can happen if agents make a circle to see each others with perfect reciprocity. Chwe (1999a) calls such communication practices "rituals" and gives a lot of empirical examples. So, it is very close to the interpretations of Foss (1999), although he does not consider a central agent (the leader).

The red hats game

A typical game where common knowledge and leadership effects are omnipresent is the game of the red hats.[24] It is a sequential game with temporal learning based on observations of rational behaviors. The game consists of three agents (children) in a room, each wearing a red or a white hat. They are asked independently, and at each period, if they know the color of their own hat. If a child gives the right answer, he leaves the room. If he gives a wrong answer, he is punished by a hard homework. If he gives no answer, he remains in the room, and waits for further questions about the color of his hat. The same question is asked every ten minutes. The children cannot communicate and cannot see the hat they are wearing, but they can see the hats their two fellow children are wearing. If one of them leaves the room, the others only observe him leaving, but they do not know the answer he gave. In this setting, nobody could leave. However, suppose now that an external agent (the teacher) tells publicly all the children that at least one

---

[21]See, e.g., Aumann (1974), Aumann (1987), Forges (1986), and Myerson (1994).

[22]The analysis of such phenomena have not been tested in experimental laboratories yet. It would be interesting to test if metaknowledge created through face-to-face communication yields significantly different behaviors than in the case of non face-to-face communication.

[23]Contrary to Dupuy (1989), who thinks that knowledge could never be perfectly transparent between agents, i.e., common knowledge, we think that such practices may do so for non-complex and non-ambiguous information.

[24]See Littelwood (1953) and Lacan (1966) for ones of the earliest versions of this example.

of them is wearing a red hat. Suppose also that they all have a red hat. Without the public announcement, no one will ever leave. On the contrary, after the announcement, they will all leave after three periods, because they will progressively learn about the color of their hat by observing others' behaviors (formal and intuitive proofs are given, e.g., in Geanakoplos (1994)). What is surprising is that, before the announce, they all know that there is at least one red hat, and they all know that they all know that there is at least one red hat. However, they do not know this last fact, i.e., it is not common knowledge that there is at least one red hat. The public announcement makes this fact common knowledge, and allows all children to leave in three periods. Thus, the teacher is a leader, as we define it before: he makes a public announcement to create common knowledge. The need for common knowledge is even larger: for the reasoning to work, we must also assume that all the children are perceptive and intelligent, hear the teacher, trust him, know the others' preferences ... and all these facts must be common knowledge. Thus, it is not sufficient that the teacher or a leader creates common knowledge, because a great deal with common knowledge must already exist in the group. Here, as in coordination games, common knowledge increases the welfare of all players.

The previous theory of leadership and of common knowledge requirement for coordination is coherent with a wide literature of the theory of games. Indeed, Ribeiro and Werlang (1989) noticed, when speaking about Nash equilibrium, the 'strong coordination requirements behind this solution concept. Not only rationality has to be common knowledge, but also the actions taken, before they are taken' (Ribeiro and Werlang (1989, p. 81)).

More recently, however, Aumann and Brandenburger (1995) showed that to achieve a Nash equilibrium, no common knowledge assumption is necessary (in a formal sense), and sufficient *epistemic conditions for Nash equilibrium* are "weaker". In fact, common knowledge of the game and of rationality is used to exclude iteratively dominated strategies and to apply the rationalizability solution concept,[25] but rationality and *mutual* knowledge of players' conjectures is sufficient to get a pure-strategy Nash equilibrium.[26] Furthermore, common knowledge is also not necessary for mixed-strategy Nash equilibrium with two players. Common knowledge is just needed to get a mixed-strategy Nash equilibrium in many players games, because conjectures must be common knowledge (see Brandenburger (1992) and Aumann and Brandenburger (1995)). Nevertheless, to justify these last results, the fact that the theory used is implicitly common knowledge remains an opened and discussed problem (see also references about counterfactual reasoning in game theory on page 18).

Note also that the analysis of leadership supposes that the leader knows just the right interaction problem, i.e., the game played, players' preferences, their available actions, and their consequences. One way to justify such leader's knowledge is to consider him as an extraneous "authority". The information of each player is shared with the leader, but remains hidden to the other players. Alternatively, the information of every player may be verifiable by the leader, but unverifiable by the players themselves. Another problem is that choosing a leader is also a coordination problem that the leader and this theory cannot solve.[27] That is why it is also useful to study how common knowledge could be created without an external agent, as we will see in section 5 when we will consider decentralized common knowledge formation. Finally, we also want to note that, even in coordination

---

[25]See Bernheim (1984) and Pearce (1984) for the initial papers on rationalizability, and Tan and Werlang (1988) for the formal relation with common knowledge.

[26]Note, however, that knowledge of others' actions is an extremely strong assumption.

[27]In the literature on correlated equilibrium, a related problem is resolved by the "super canonical device" (Forges (1986)), which is a solution that insures that there is no point worrying about the choice of a communication device.

games, a leader might solve incentive conflicts in asymmetric coordination games. Indeed, consider the "battle of sexes" game of figure 2.

		Player 2	
		A	B
Player 1	a	$(2,1)$	$(0,0)$
	b	$(0,0)$	$(1,2)$

Figure 2: A "battle of sexes" game.

Obviously, player 1 prefers the Nash equilibrium (a, A), rather than the Nash equilibrium (b, B), and the contrary for player 2. If the game is repeated, a fair leader might make (a, A) and (b, B) common knowledge alternatively. But, in this context, there is a priori no reason for the agents to follow the leader's advice.

Another tool for common knowledge creation is advertising, which will be analyzed in the next subsection.

2.3   Advertisement

In this subsection, based on works of Chwe (1998a, 1998b), but related to the previous analysis, we show how some kinds of *advertisements*, especially those for social goods, could solve some coordination problems, once again by creating common knowledge.

We remind that *social goods* are goods that are interesting to buy only if many other agents buy it, because of networks externalities, technological compatibility, for social reasons or to conform to community standards. Such goods include soft drinks, beers, movies, clothing, computers, softwares and computer operating systems upgrading, large parties or musical events, communication products, fax machines, financial services ... In this case, the seller of the product tends to advertise the product publicly, for example on popular and extensive television shows or newspapers. Other examples include advertising where agents are informed publicly and simultaneously, as in large advertisements in sports meetings ...[28] With such advertisements, potential buyers are not only informed by the information *contained* in the advertisement, but also that many others have seen this advertisement, knowing that many others have seen it, and so on. Thus, in advertisements for social goods, not only the message content and meaning (information about quality, innovation or differentiation) matters, but also its openness and publicity characteristics, i.e., metainformation. This is very close to the view of leadership: leadership could align incentives, but it could also influence beliefs and create common knowledge to help coordination. Social goods advertisements, by their content, could influence potential buyers' evaluation of the product, and by its publicity it creates common knowledge of the product and helps coordination.

Chwe (1998a) illustrates his theoretical idea by an empirical analysis showing that, for social goods, audience size and advertising costs are usually higher than for non-social goods. More precisely, taking 119 brands in the US during three months of a television season in 1988–1989, he shows that advertisers of social goods often advertise their products on large audience television shows and spend consistently more per viewer than advertisers of other goods. One could explain such heavy ads by the fact that they constitute some proofs of a high quality product because, if not, advertisers might not spend too much money for their products. Another explanation would be that such

---

[28]Chwe (1998b) gives the example of Apple which made a big advertising during the Super Bowl in January 87.

ads might easily change customers' preferences. Nevertheless, this does not explain why advertisements of social goods tend to be done in more popular shows. Hence, common knowledge is one of the reasons. This confirms the previous idea, i.e., that advertisement has publicity effects. As a result, the impact of advertising by telephone or mail tracts distributions would not have the same effect as mass audience advertisement, even if the audience size is the same.

We should note that the need for a common knowledge basis for coordination is also emphasized in some papers on organizational learning, even if the definition of common knowledge is less demanding than the definition of game theory. In contrast, the definition of "knowledge" is more general than in game theory (see Marengo (1992) and Crémer (1990, 1993)).

2.4   Weakening common knowledge and coordination conditions

We have seen that common knowledge is necessary to solve coordination problems. This could be formally shown, e.g., in the electronic mail game. Such common knowledge could be created relatively easily by advertisement, leadership and public reports.

In this subsection we show how to solve coordination in incomplete information settings, even with possible noisy communication. In subsection 2.1 we explained that if the communication system is imperfect (noisy) and if a message is sent at each period in accordance with the reception of a message, strategic coordination is impossible. This was actually shown by Rubinstein (1989). Computer scientists already showed that with such a communication system, whatever the rule of action considered, perfect coordination is impossible (see Halpern and Moses (1990)). Thus, common knowledge seems to be indispensable for coordination. However, one can solve this problem in different manners.

First, one can have a Bayesian maximization approach so that perfect coordination (successful coordination with probability one) is not necessary if agents are not strategic, i.e., if we impose some rules of actions on them (as it could be done in computer science). In this case, if the communication system is sufficiently reliable, there exists a rule of action such that the coordinated attack happens almost each time the enemy is not prepared. The rule of actions consists for player 1 to send a message and to attack immediately iff the enemy is unprepared, without waiting any confirmation, and for player 2 to attack if he receives a message. This is possible, because with a Bayesian view and with agents maximizing their utility, the gains (or losses) which happen with small probability are usually not relevant. But with strategic and decentralized behaviors it is not the case, as we saw before. Thus, such a solution is not acceptable. Agents act in a short-sighted way since they make no allowance for the reception of possible confirmations (in particular, agent 1 attacks without taking into account the reception of a confirmation). This is conceivable if the generals received preliminary orders imposing this rule of action.

Another solution is to consider a simpler communication protocol, where messages are not sent back and forth, which is in fact a naive protocol, but a communication system where just one message is sent, from the first player to the second. In this case, if the communication system is sufficiently reliable, the simple protocol makes it possible to have almost always a *strategic* coordination.[29] Thus, common knowledge is not necessary.

The preceding communication system seems to be arbitrary, because imposed on agents. However, by considering strategic coordination *and* strategic communication at the same time, we can prove that such a protocol could be generated endogenously. Indeed, Binmore and Samuelson (1999) show that if communication is strategic, then a Nash

---

[29]In fact, the simple protocol with strategic agents is equivalent to the previous naive protocol with non-strategic agents.

equilibrium where just one message is sent exists. It should be noted that to reach this solution, a lot of common knowledge assumptions are implicitly supposed. In fact, common knowledge that the second general is not informed whether the enemy is unprepared, of the rules of the game, of the preferences over the outcomes, of the communication coding ... is useful for efficient communication and behaviors.

To explain the differences in behaviors between the "naive" communication protocol and the simple one, we must look at the beliefs generated in each case. As a matter of fact, common belief that the enemy is not prepared to fight is never obtained in the "naive" communication protocol, contrary to the simple protocol. More precisely, common belief is as common knowledge by replacing knowledge by belief with a certain probability.[30] In the simple communication protocol, we get a $(1 - \varepsilon)$-common belief, when it is impossible to find a belief probability superior to $1/2$ where it is common belief that the enemy is unprepared with the "naive" communication protocol (see Morris and Shin (1997)). Nevertheless, if coordination is on a continuum of strategies, common belief is never sufficient, and strict common knowledge is required (see Shin (1996)).

Afterwards, to solve the coordinated attack problem, one can consider limited rationality. This was done by Dulleck (1997) who introduced *imperfect recall*. This allowed him to change agents' induction reasoning by modifying a little the information structure. Suppose that a player is not able to distinguish, after T messages received, if he gets T, $T + 1$, ... or $T + l$ messages. The intuition behind this modeling is that this agent is not able to interpret the messages after he gets T messages (as we can hardly distinguish orders of mutual knowledge superior to 3 or 4 levels). If $l \to +\infty$ we can interpret this modification as the fact that he does not take into account more than T messages. The interpretation in terms of imperfect recall is that he forgets if he received $T, T + 1, \ldots$ or $T + l$ messages. With such a modeling, there exists a Nash equilibrium where coordination happens if each player receives at least T messages.

Finally, if strategic communication is not allowed, another solution was found in computer science. When there is temporal imprecision (which is almost equivalent to noisy communication), common knowledge is not attainable, and thus perfect coordination is impossible. To solve this paradox, Fagin, Halpern, Moses, and Vardi (1999) proposed to modify the *granularity of time* (i.e., to model time in a less precise manner to get simultaneous communication) or to weaken coordination conditions. With temporal imprecision we speak about "ε-coordination", and with noisy communication we speak about "eventual coordination", which corresponds in fact to the probabilistic (Bayesian) view of the problem discussed at the beginning of this subsection.

3   Negative effects of common knowledge

In the preceding section we have seen that coordination and common knowledge are related phenomena, and particularly that common knowledge is essential for coordination and for welfare improvement. Nevertheless, we will see in this section that common knowledge of various things (of the interaction problem, of preferences, of rationality, of intentions) could be collectively bad, when first-order or low-order knowledge could make agents better off.

---

[30]Practically, common belief is obtained when there are "almost" public events, i.e., events that agents believe with high probability whenever they occur.

3.1 Cooperation and conflicts of interest

Common knowledge assumptions often have bad effects on welfare even when few conflicts of interest are present. These strong assumptions also explain the non-intuitive and unrealistic results we sometimes obtain in game theory. To illustrate the harmful effects of common knowledge we will essentially focus on two kinds of games: cooperation games (through the prisoner's dilemma), and centipede like games. The first kind of game is well known and represents certainly, with coordination problems, one of the most common economical and social problems. The latter is more particular, but it is in this particularity that we find all its relevance. In such games, it is essentially on the inductive and iterative reasoning of agents that common knowledge is harmful, because it eliminates Pareto optimal outcomes.

The prisoner's dilemma

The well known history of the prisoner's dilemma is the following. Two suspects in a crime are put into separate cells to be questioned about their culpability. If they both confess (denounce), each will be sentenced to three-year imprisonment. If only one of them confesses, he will be freed and the other will be sentenced to four years. If they both cooperate they will only be sentenced to one year.[31]

By associating utilities to the number of years in prison we can get the normal form game of figure 3. Actions C and D correspond respectively to cooperation and denunciation.

$$
\begin{array}{cc}
 & \text{Player 2} \\
 & \begin{array}{cc} C & D \end{array} \\
\text{Player 1} \begin{array}{c} C \\ D \end{array} & \begin{array}{|c|c|} \hline (0,0) & (-5,1) \\ \hline (1,-5) & (-1,-1) \\ \hline \end{array}
\end{array}
$$

Figure 3: A prisoner's dilemma.

It appears clearly that cooperation is strictly dominated by denunciation. Thus, the only Nash outcome is (D, D), which is Pareto dominated by (C, C). The dilemma is that individual rationality leads agents to denounce when mutual cooperation would give them a larger utility. The dilemma is a paradox when we consider a repeated but finite prisoner's dilemma game. Indeed, suppose that we repeat it T times. In this case, if the game (including preferences), the number of repetitions, and rationality are common knowledge, iterative elimination of (weakly) dominated strategies leads agents to denounce at each period. Backward induction gives the same result. Thus, it constitutes the only subgame perfect equilibrium. It constitutes also the only Nash equilibrium outcome.[32] This paradox will be solved by relaxing common knowledge assumptions (but by keeping lower-order knowledge assumptions). First, we present the centipede game, where the same kind of paradox appears.

The centipede game

Two players are around a table with ten pennies on it. At each period they could alternatively take one or two pennies (these actions are respectively noted C and S). If just

---

[31]Examples of such a game include all free rider problems such as those of public goods and international cooperation.

[32]If the number of repetitions is infinite, cooperation constitutes also a subgame perfect equilibrium. But in this case, all outcomes are actually possible (by the folk theorem).

one penny is taken, the game continues. If a player takes two pennies at once, the game stops. If they always take just one penny, the game ends when no penny remains on the table. This game is a version of the centipede game introduced by Rosenthal (1981), also called "penny pinching" game. An extensive form representation is given in figure 4.
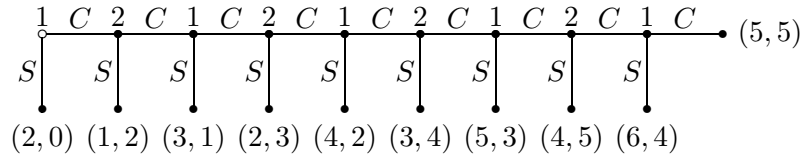

Figure 4: The centipede game.

As in the case of the prisoner's dilemma, there is only one subgame perfect equilibrium. It consists for the players in taking two pennies at each round. Thus, the first player gets two pennies and the second gets nothing. This is independent of the number of pennies on the table: when several millions of pennies are available, the first player will only get two of them, and the second nothing. This situation is clearly Pareto dominated by almost all outcomes. Again, common knowledge of the game and of rationality leads to a very inefficient outcome.

Efficient behaviors without common knowledge

The preceding paradox cannot be resolved with standard assumptions. Indeed, common knowledge of rationality and of the game played imply *backward induction*, and thus a subgame perfect equilibrium.[33] However, such a result is not confirmed in experiments and is not compatible with our intuition. Moreover, the subgame perfect solution is not only a paradox because it is counterintuitive and contradicts reality, but it constitutes also a *theoretical paradox*.[34] An internal logic problem does appear in the backward induction process: agents begin to reason at the end of the game when they know that, by the solution, they will never be at such nodes. Their reasoning at these nodes would cause them to revise their theory, but it is not the case. Insofar as nothing could be said about what would happen off the backward induction path, it is difficult to support a theory claiming that a deviation is always irrational. This paradox is in fact due to the appreciation of counterfactual and hypothetical reasoning in game theory (a counterfactual statement is a statement about a world commonly known to be impossible). For more discussions and debates see, e.g., Binmore (1996), Aumann (1996, 1998), Samet (1996, 1997), Gilboa (1998), and Halpern (1998).

A whole host of objections to the backward induction solution have occurred to many working game theorists. Consequently, many solutions aiming to explain cooperation were given in the literature. We can actually solve the problem by introducing automata, altruism, sequential reciprocity or generosity, imperfect recall, imitation, by preference manipulations, by a benevolent authority control, by legal and social sanctions ... However, these models alone do not really account for the paradox. Another solution might

---

[33]The fact that common knowledge of rationality implies backward induction in perfect information games was formally shown by Aumann (1995). Some weakness of Aumann's (1995) model have been criticized and modified by Stalnaker (1996) and Halpern (1998). They are related to the fact that Aumann (1995) does not allow players to revise (update) their beliefs about rationality during the game.

[34]The same kind of paradox appears in the chain store game (see Selten (1978)), which is a finite repetition of an entry game. A solution, also related to common knowledge, was given by Milgrom and Roberts (1982).

be to consider "trembling hands", i.e., to explain some deviations by small random errors. Presumably, such a solution is not acceptable if we keep all other assumptions because we know that common knowledge of rationality implies rationality which forbids, by definition, any kind of mistakes in decision making. We think that the best way to explain and to resolve the paradox is to slightly relax the strong common knowledge assumptions.

Pettit and Sugden (1989) analyze the finitely repeated prisoner's dilemma by keeping all standard assumptions, except common knowledge of rationality, which is replaced by *common belief* of rationality: everybody believes that everybody believes ... that everybody is rational.[35] In this case, cooperation is possible, especially with the "Tit For Tat" strategy, because common belief allows false beliefs and beliefs reconsideration. This strategy, which consists in cooperating first and then in cooperating at each round iff the other player cooperated in the previous round, proves to be the most successful in the well known tournament of Axelrod (1984). In a more formal model, Reny (1993) also establishes that common belief of rationality should not imply the defect-always outcome. The repeated prisoner's dilemma with rational cooperation but without common knowledge of rationality was also analyzed by Kreps, Milgrom, Roberts, and Wilson (1982).

Cooperation is not only possible by relaxing the common knowledge of rationality assumption. Indeed, one can introduce incomplete information about the number of repetitions, or about the number of pennies on the table. In this case, even if it is common knowledge that this number is finite, but the exact number is not, the backward induction solution can be broken down.[36] Relaxing common knowledge of preferences is another solution: if some uncertainty about players' payoffs is introduced, Pareto superior outcomes could be obtained, even with common knowledge of rationality and of the length of the game.

3.2 Negotiation and trade

An interesting phenomenon, highlighted by Ayres and Nalebuff (1998), is that during *negotiations*, agents who communicate often want to reveal only first-order information, but not higher-order one, because it may be inefficient for negotiation, even if first-order information may help agreement. To solve such problems, they propose to use a mediator who talks to each agent *independently* (non publicly) so that only first-order information is transmitted between agents, without interactive and common knowledge creation. Thus, they see the mediator exactly in an opposite sense compared to section 2. Here, the mediator makes *secret* information revelations. He could be seen as a lawyer, a negotiation consultant or a friend. What is revealed is common knowledge between each agent and him, but not between the agents. Then, mediation and leadership may create common knowledge in coordination problems, but could also solve other problems such as negotiation by, on the contrary, preventing the creation of common knowledge.[37]

To show that such phenomena exist in negotiation problems, i.e., that first-order knowledge can sometimes facilitate agreement, but not higher-order knowledge, let us present some examples. To begin with, consider a wife and a husband who are having troubles in their relationship, and suppose that the husband wants to change things to improve their relationship. He wants to make clear to his wife that she has to change her behavior partly. The communication could work as follows:

Husband: 'Honey, I want this relationship to work but if you can't make these changes then I want a divorce';

---

[35]See Monderer and Samet (1989) for a formal analysis of common belief.

[36]See Binmore (1992, pp. 479–481).

[37]However, note that practically, a mediator cannot always prevent higher-order information transmission, as he cannot always create strict common knowledge in public communication.

Wife: 'You got it';
Husband: 'I didn't want the divorce, I wanted the changes'.

At this point, the threat is common knowledge. Not only the wife is informed by the fact that if she does not change her behavior, a divorce is possible. She knows that her husband thinks about a divorce. Thus, she would be reluctant to stay in marriage with someone who has such ideas. Moreover, he knows she knows his threat, so he knows she could be reluctant to stay in marriage, and this may make their relation difficult. Instead, if a marriage counselor tells the wife what the husband started to say, the behavior might be different, because in this case she would not know that her husband thinks about a divorce and she might effectively change her behavior. Such effects are difficult to formalize, but they clearly matter.

The same situation could happen in a buyer and a seller partnership. The seller may say:

'Look, we are the only company in a position to produce the goods right now for you, so if you want your supply, here's the price and you better sign this contract or your factory is going to get shut down'.

If the buyer was only informed by the content of this message, he might sign the contract because he has few alternatives. But common knowledge of the threat may cause their relationship to break down as in the divorce example. Here, the buyer knows that the seller might be opportunistic, and so he might be less cooperative. The seller also knows this, and might not have any confidence in their partnership too. Knowing this, the buyer might trust the seller even less, and so forth. The mutual lack of trust being common knowledge, trust is itself less and less effective. A mediator may, once again, help to avoid such learning just by informing the buyer that the supplier might break off relation if he does not pay the higher price. Ambiguous statements might also solve the problem. As a matter of fact, the seller could say directly to the buyer: 'Think about what would you do if you were in my shoes'. In this case, the buyer could acquire the preceding information (that he does not have a negotiating power, and that he has to accept higher prices) without the seller knowing that he knows and without the buyer knowing exactly what the seller thinks. We see that in these kinds of situations, ambiguous statements can be preferable and welfare improving.

Ayres and Nalebuff (1998) describe similar examples, including organizational relationships, where another way to transmit only first-order information is to make anonymous notes, or to provide suggestion boxes. Another final example could concern a relationship between a parent and her teenager. If she (the parent) learns he (the teenager) has tried marijuana in her presence, this would be common knowledge and would generate inefficient behaviors. Knowing that she knows, the teenager might use marijuana more secretly. The parent, knowing he knows she knows would be forced to punish him, and maybe to react in an intolerant and inefficient way. To solve this problem she might ask a friend, a member of the family, or an instructor to inform him that what he did is bad for himself.

In a game-theoretic setting, such a phenomenon was actually proved formally by Neyman (1991), who showed that information can never hurt an agent, if the others are not aware he has it. This can happen, e.g., if an agent secretly acquires some information. Thus, what may hurt an agent is not his knowledge improvement (about "fundamentals" or about others' knowledge) but the fact that other agents know that his knowledge has been improved. The general claim that information may hurt in game theory is due to the fact that, when we give an agent more information in a game tree, the others will know that he has additional information, because the complete game (including information asymmetries and the move of the "Nature") is common knowledge.

4 Asymmetric common knowledge effects

In the preceding section, we have seen how common knowledge is, again, very influential on interactive behaviors. But, contrary to the problems studied in section 2, common knowledge induced suboptimal behaviors. Now, we are going to see some examples where common knowledge and interactive knowledge also imply *asymmetric* effects. Even if everybody has the same first-order knowledge, the configuration of what everybody knows about others' knowledge may make some agents better off, and others worse off.

We first present a very simple bargaining game, but with a rich information structure, introduced by Lipman (1990). Afterwards, we consider a negotiation game where different levels of interactive knowledge might modify the negotiation powers of agents in a monotonic manner. This situation was analyzed by Ayres and Nalebuff (1998), and we formalize it as a game close to the centipede game. Finally, we briefly present some relations between speculation, or betting, and common knowledge.

4.1 A simple bargaining game

Consider a bargaining game where a seller of a single indivisible good proposes a price to a buyer, who can accept or reject it. If he accepts, they trade at the proposed price. The seller gets a payoff equal to the price minus the production cost and the buyer gets a payoff equal to his valuation minus the price. For trade efficiency, this valuation is always supposed to be superior to the cost. Otherwise, if the buyer rejects the seller's price, they both get a payoff of zero.

- If the preceding game description and parameters are common knowledge, the seller trivially sets a price equal to the buyer's valuation, who gets a payoff of zero at equilibrium.

Now suppose that the seller's cost and the buyer's valuation are determined by a state of the world, belonging to a large but finite state space. Lipman (1990) proved the following:[38]

- If the buyer does not know the true state of the world, and this is common knowledge, the seller sets a price such that the buyer's expected payoff is zero.

Thus, if what the buyer knows is common knowledge, the seller gets all the bargaining power and the buyer gets an expected payoff of zero, whenever he knows his valuation or not. If what the buyer knows is not common knowledge, the result is quite different:

- When the seller is sufficiently uncertain about the buyer's knowledge, i.e., if the seller believes the buyer is uninformed with a probability not close to one or zero, the buyer's expected payoff may be strictly positive.

Thus, in this problem, whether the seller knows or not if the buyer knows some parameters matters in equilibrium outcomes. If the buyer can choose, after observing the seller's choice, to learn about the state of the world, we get very surprising results. Indeed:

- If the buyer can choose to freely acquire the information about the true state, he will acquire it, and we get endogenously the same situation as in the complete information game, where all is common knowledge.

---

[38]The solution concept used was the perfect sequential equilibrium.

- Nevertheless, for any positive, even very small cost of information acquisition, the buyer's expected payoff is strictly positive, and not to acquire information is no more a dominated strategy.

Intuitively, the reason for such discontinuity in the cost of information acquisition is the following. Since this cost is strictly positive, the seller does not set the price at the buyer's valuation, because if he does so, the buyer would not pay for information, since this valuation is revealed by the price. Thus, the seller would have an incentive to fix the highest possible price, which would no more reveal the valuation. Then, such solution is not stable.

We see that when the buyer's knowledge is common knowledge, he gets no gain from trade. However, when his knowledge or ignorance is not known by the seller, the buyer gets a surplus. This happens with a corresponding exogenous information structure, or with an endogenous information structure with a strictly positive cost of information acquisition.

Then, if the valuation of the good is perfectly transparent and commonly accessible, the buyer may acquire such knowledge perfectly and freely, the seller might know this, and the buyer's payoff would be zero. Such perfect knowledge accessibility will give a complete bargaining power to the seller. He is the potential "winner" of the game; when this one is common knowledge, he derives his strategic advantage from it.

4.2   Labor-management negotiation

Consider a labor-management negotiation, occurring at the beginning and possibly during a strike. The problem is to know which part will concede and accept the other's position. Denote L for the labor and M for the management. Suppose that the cost of management if it gives in to the union's demand is $B_M = 100\$$, and the benefit of labor, i.e., the present value of the job security it has demanded, is $B_L = 60\$$. We have $B_L < B_M$, i.e., job security might cost management more than it benefits labor, because it makes production less efficient.

Each day, each party can choose to concede (action C) or not to concede (action $\bar{C}$). Without loss of generality, suppose that labor is the first to choose at each period (day). During the strike, each party incurs some cost. We suppose that labor (management) cost per day is 1\$ (3\$). They are respectively endowed with a fund $F_L = 70\$$ and $F_M = 300\$$. When a party's funding runs out we suppose that it must concede. Thus, labor will concede before the 71th day, and management before the 101th day. We know that after day 70, the only possible play is for labor to concede, with a payoff of zero, management getting a payoff of $300 - 70 \times 3 = 90$. The game ends at this period. The extensive representation is given in figure 5 on the next page (the first coordinate (respectively second) of the payoffs corresponds to the payoffs of labor (respectively management). We have respectively represented the days 1, 2, 3, 4, 5, 36, 37, 38, 69 and 70.

In order to explicit the agents' decision procedure, we consider (weakly) dominated strategies as a solution concept. Suppose that the game is common knowledge, except F_L. If management knows F_L, it will know that labor concedes at day 70. Thus, it does not concede, because it gets 90\$ instead of -27\$. From day 38 to the end, it gets less than 90\$ if it concedes, so it will hold out from day 38. Now, if labor knows management knows B_L, it will know that management hold out from day 38. If it concedes after, it gets less than 33\$ for sure, thus labor concedes at day 38 to get 33\$. If management knows this, i.e., knows labor knows management knows B_L, it will know that if it does not concede after day 5, it will get at least 189\$. Thus, management holds out from day 5 (if it concedes at day 4 it could get 191\$ so we cannot conclude more from here). If labor knows this, i.e., labor knows management knows labor knows management knows B_L, labor will concede at day
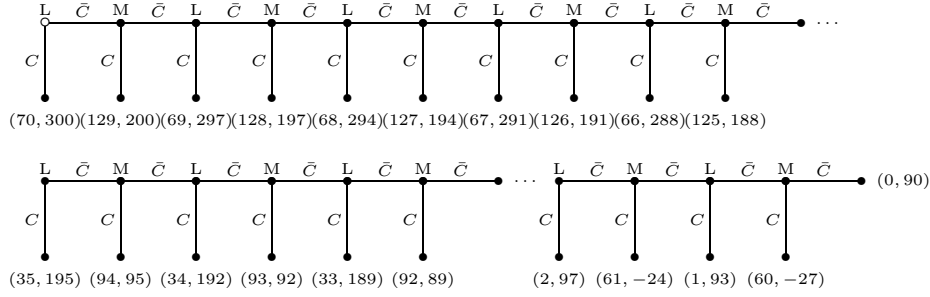
Figure 5: A labor-management negotiation.

5 with the same reasoning as before. Before day 5 we cannot conclude because labor does not know if management concedes or not before day 5, so it could be in its interest not to concede before day 5, because labor could for example get 126\$ if it expects management to concede at day 4. Knowing this, i.e., knowing that labor knows management knows labor knows management knows B_L, management will hold out from the first day because so, it is sure to get at least 288\$. Finally, knowing this, i.e., knowing that management knows labor knows management knows labor knows management knows B_L, labor will concede at the first day, because it cannot get more than 70\$ if it does not concede from here.

We can see that the length of the strike and the negotiation power of management depends crucially on interactive knowledge. To be sure that labor concedes the first day, and thus that management uses all its bargaining power, one needs at least six levels of interactive knowledge. We can notice that if all is common knowledge, except rationality, the same reasoning applies by analyzing the game for different hierarchies of knowledge about rationality. Note also that the resolution of this game is as paradoxical as the resolution of the centipede game or the repeated prisoner's dilemma. Indeed, suppose that you represent labor, and that we assume that all is common knowledge. Being rational imposes that you concede the first day. But suppose that you do not concede. In this case, management could not trust the common knowledge assumptions and you may force him to concede ... This question leads to ask the seemingly stupid question: "Is it rational to be rational?"! Such a problem remains a very uncomfortable one for game theory (see references we gave on page 18 about the backward induction paradox and about counterfactual reasoning).

4.3   Speculation and betting

A general problem where differences in higher-order knowledge about intentions or action might asymmetrically affect behaviors and utilities is *zero-sum betting* or gambling. Consider the following example, taken from Sebenius and Geanakoplos (1983). Two players want to bet about the realization of an event A. The information structure is represented in figure 6 on the next page. The big rectangle is the state space, with an uniform probability distribution, and (p_1, p_2, p_3), (q_1, q_2, q_3) are the partitions of players 1 and 2, respectively. Each player knows in which of his partition the true state ω is. We consider the situation where ω belongs to A, p_2 and q_3. The bet works as follows. If $\omega \in A$, player 2 gives one dollar to player 1. If not, player 1 gives one dollar to player 2. Agents are assumed to be risk neutral so that they maximize their expected gain. player 1 (2) will accept (reject) the bet if he believes with probability superior to half that A realized,

given his private information, i.e., given his reached information set. We suppose that, at each period, players successively announce their willingness or unwillingness to take the bet.
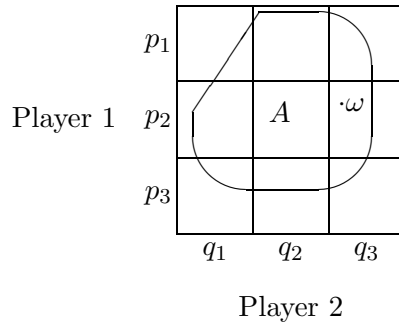

Player 2

Figure 6: Betting with private information.

Suppose that player 1 is the first to announce his intention to bet. Since he knows that $\omega \in p_2$, he believes that A realized with a probability superior to half, and accepts the bet. In that case, player 2 learns that $\omega \notin p_3$, because if not, player 1 would not have chosen to bet. The conditional probability for player 2 that $\omega \in A$ given that $\omega \in q_3$ and $\omega \notin p_3$ is less than half, and also wants to bet. player 1 infers from this willingness to bet that $\omega \notin q_2$. player 1's belief about A is equal to the conditional probability of A given that $\omega \in p_2$ and $\omega \notin q_2$, which is still more than half, and he continues to accept the bet. Thus, player 2 learns that $\omega \notin p_1$, i.e., knows that $\omega \in p_2$ and continues to accept the bet. Then, player 1 deduces that $\omega \in q_3$ and cannot, now, accept the bet, because the conditional probability of A given that $\omega \in p_2$ and $\omega \in q_3$ is less than half. We see that the bet was accepted during four periods, but player 1 rejects it at the fifth period. Then, if the willingness to bet is known by each player, a bet is possible. In this example, player 1 would actually win it. However, if players communicate back and forth their willingness to bet, one of them will reject it in a finite number of periods. Sebenius and Geanakoplos (1983) proved that this phenomenon is always observed with the standard information structure presented in section 1, i.e., that any bet is always rejected if agents communicate their intention back and forth. That is, it cannot be common knowledge that they both want to bet because it cannot be common knowledge that they both expect to win, but it could be mutually known.[39] Thus, mutual knowledge could make an agent better off, and the other worse off, but common knowledge of each one's gain expectation forbids any bet or gamble; possible remaining information asymmetries are not relevant, although higher-order information asymmetries might be, before the creation of common knowledge through willingness revelations.

A related result about *speculation* was found by Milgrom and Stockey (1982) for the theory of markets with asymmetric information. Suppose that it is common knowledge when a trade takes place that it is feasible and individually rational. They showed that if an allocation is Pareto optimal, and agents receive private information but share the same prior beliefs, they will have no incentive to trade, even with this additional information. The intuition of this no-trade result is that, given that the allocation is initially Pareto optimal, any incentive to trade will be seen as an opportunistic behavior, as a motive to bet. Such willingness being common knowledge at equilibrium, the trade cannot be acceptable for all traders, for a similar reason that they cannot bet in the preceding example: if someone is willing to trade, someone else will deduce that the trade is unfavorable for

---

[39]This results comes form the "agreeing to disagree" literature that we will briefly evoke in section 5.2.

him. With mutual or higher-order knowledge, but without common knowledge, a non-null trade is however possible, and some agents may take some advantage from their private information. Neeman (1996) even shows that replacing common knowledge of rationality by common (even strong) belief of rationality allows for arbitrary large volumes of trade.

5 Decentralized knowledge evolution and communication

We have seen that public events or revelations are obvious and efficient tools to manipulate beliefs and to create common knowledge. Nevertheless, such a creation is independent of the interacting agents, who do not control it, and an external agent is not always present. In this case, agents may communicate among themselves to distribute their knowledge and modify the information structure. In this section, we will describe how knowledge and common knowledge could evolve in such a *decentralized* manner, i.e., without an external agent such as a leader, an advertiser or a mediator. We will call *knowledge certifiability* the possibility to transcript perfectly (with full preciseness and trust) knowledge into messages that include certified information, through a given and common language.[40]

We begin by studying common knowledge creation in network communications, with non-strategic information revelation and with perfectly certifiable knowledge. Afterwards, we analyze knowledge and interactive knowledge evolution when certifiability is not perfect, i.e., all knowledge cannot perfectly be revealed by messages. We also see how common knowledge sometimes helps for *agreement* and *consensus*. Finally, we consider *strategic communication*. The analysis of endogenous knowledge evolution is still very limited in the literature, and we will emphasize the need for theoretical studies of voluntary information structure's manipulations in many decision makers' interactions with interactive knowledge.

5.1 Evolution of knowledge in networks with perfect certifiability

Chwe (1999b, 2000) studied efficient interaction structures to solve coordination problems by analyzing common knowledge evolution in network communications.[41] He focus on a particular coordination game with incomplete information where, to have some incentives to participate in a collective action, it is not necessary that all agents participate. To participate, an agent only needs that a certain number of agents participate. This number is called the "threshold", and may be different for each agent, and is not common knowledge. This threshold could for example be a parameter of risk aversion of an agent (if the threshold increases, it means that the agent is more prudent before engaging in the collective action). In a standardization context, it could correspond to a technology parameter, a market power, or to the size of a firm. In a problem of collective revolt, a threshold could correspond to the integrity of an agent. Finally, in a coordinated attack problem, it could correspond to the power of an armed division.

The game is symmetric in the sense that there is no disagreement on what to coordinate. At each period, each player reveals non-strategically his threshold to players who are directly related to him by the communication network (represented by a deterministic

---

[40]This definition of knowledge certification is not so far from the definition of knowledge codification given by Cowan and Foray (1997), who define codification as a process of knowledge conversion into messages that could be manipulated as information.

[41]The network is not a network which diffuses behaviors, like in evolutionary models of contagion or of local interactions where neighborhood actions may directly influence decision making. A communication network is just a network where information and metainformation are transmitted. However, this does not mean that it does not indirectly and dynamically influence individual behaviors.

graph). At each of these periods, each agent chooses to participate or not in the collective action, given his new acquired knowledge. The communication system is supposed perfectly reliable, contrary to the communication in the electronic mail game.

Once the interaction problem defined, Chwe (1999b, 2000) shows that participation grows over time and with the network density, and decreases with the players' threshold. Such growth of participation is done through the evolution of common knowledge about the thresholds: if it is common knowledge in a group that the thresholds permit participation, players in the group will engage in the collective action. Again, common knowledge plays a crucial role, because inferior levels of interactive knowledge (such as mutual knowledge) may be insufficient. For example, if four agents know that they all have a threshold inferior to four, this would be enough to coordinate, but it is not the case with decentralized behaviors: the four agents have also to know that they all know that they all have a threshold inferior to four, and so on. Achieving such knowledge may take a lot of time, so it is sometimes cheaper to use a leader to help coordination, as in subsection 2.2.

However, we can find some efficient networks for common knowledge creation. With *strong links* (i.e., when agents directly communicate with the neighborhood of their neighborhood, i.e., links are very close) common knowledge is created rapidly for low thresholds, and slowly for high thresholds. With *weak links* (i.e., when agents do not communicate directly with the neighborhood of their neighborhood, i.e., links are very dispersed) it is the contrary. Strong links are efficient to create common knowledge at a local level, and thus to generate coordination of small groups where the thresholds are low. The reason is that weak links spread information quickly, when strong links traverse large groups slowly, but is more efficient to distribute higher-order information, and thus to create common knowledge locally.

It is clear that if all agents are directly linked, i.e., if the corresponding graph is complete, we get the best network for such a game (where more information and metainformation is always better). Such networks correspond to public communication, and so common knowledge is created in just one period (remember that knowledge is perfectly certifiable and there is no communication noise). For non-complete graphs (with a limited number of links), the best networks are generally "star" shaped networks, i.e., centralized networks. The worse networks are those forming a line. But it is not a general rule: the quality of a network depends on the agents' thresholds. The "star" network will prove to be influential in the next subsection too.

In this subsection, we have implicitly supposed that all agents' knowledge (and metaknowledge) can be transmitted. Let us see what happens when not all knowledge can be transcript into messages.

5.2 Communication process with imperfect certifiability

A part of the common knowledge literature is devoted to the analysis of conditions we have to impose on sent messages to get common knowledge of them, and then a consensus.[42] In some sense, these conditions could be interpreted by knowledge certification possibilities. These works start with Aumann (1976), who proved that agents cannot "agree to disagree", i.e., cannot disagree about their posteriors probabilities if they are common knowledge. In other words, rational economic agents, having common knowledge of their beliefs about an event, may not disagree. Thus, common knowledge is a condition for

---

[42]A consensus in the sense that all sent messages become the same. If all agents have the same decision rule, i.e., are "like minded", messages could also be viewed as observable actions. With such an interpretation, works presented here are directly related to betting and speculation results of subsection 4.3 on page 23.

agreement.

Later, several authors were interested in the way beliefs may become common knowledge. Their main result is the general convergence of sent messages when communication takes place (but it is not strategic at all). Their convergence to a unique message depends on the available communication system. First of all, Geanakoplos and Polemarchakis (1982) showed that if two agents exchange their posteriors about an event, the revision of them converges in a finite number of periods (if the information structure is finite) and these probabilities become common knowledge, and then equal. Roughly speaking, "we cannot disagree forever". Afterwards, Cave (1983) generalized this result to public communication with many agents, when sent messages verify the sure thing principle, i.e., the function which gives the information sent is union consistent. Finally, Parikh and Krasucki (1990) showed that if communication is not public, but pairwise, i.e., through a communication network, the union consistency condition is not strong enough. Actually, they showed that sent messages must verify a convexity condition, and the graph of communication must be strongly connected. Thus, for common knowledge and consensus to be achieved, the conditions on sent messages are stronger: in some sense, certifiability (the available language) must be richer. Krasucki (1996) showed however that with a non-cyclical graph (including the "star" types of communication protocols), union consistency is sufficient. Hence, certifiability conditions needed for learning, common knowledge, and consensus are weaker, depending on the type of communication practice and interaction.

All these studies are sometimes unsatisfactory because information transmission is *mandatory*. To understand endogenous knowledge evolution through decentralized communication, and the role of certifiability, we must consider strategic information revelation. Some ideas are briefly presented in the following and last subsection.

5.3 Strategic knowledge certification and endogenous information structure

Under efficient conditions of communication and certifiability, knowledge could be homogenized (partially or completely), even if this process may take some time. Such learning processes influence equilibrium behaviors, but also collective agreements, consensus and beliefs convergence. Nevertheless, in various interactive situations, knowledge sharing is clearly not always observed even under favorable conditions of communication. Agents who share their knowledge do it as they choose some specified actions, i.e., they do it strategically.

Strategic and decentralized communication could be seen in two ways. One can consider information revelation when agents can lie. Such games, with preplay and cheap communication are called *cheap talk games* which are particular signaling games where the signals do not affect directly the payoffs. Such communication is restrictive because often only one agent could send messages and is the only one having private information. Then, there are no heterogeneous interactive knowledge. Another literature deals with strategic information revelation with provable and certifiable information. Certification possibilities are exogenous but the use of them (i.e., disclosure of certified information) is endogenous because determined by the independent and strategic choices of the agents. Such literature focus on sufficient conditions on the game and on certification possibilities to get perfectly revealing equilibrium. Yet, interactive knowledge is not considered, and one of the most general model was proposed by Okuno-Fujiwara, Postlewaite, and Suzumura (1990) where each agent may communicate publicly, simultaneously, and may have private information. The other works in this literature are almost exclusively applied to *persuasion games*. It started with Grossman (1981) and Milgrom (1981b) who analyzed

close models of persuasion where a buyer purchases some quantity of a product having an unknown quality. The seller, who has some private information about the quality, can reveal or conceal some of it at no cost and cannot make false reports. In such a setting, it was shown that the seller is not able to mislead the potential buyer about the quality of his product. Thus, if all information can be certified by the seller, the equilibrium decision is the full-information decision. Subsequent papers generalized this result by considering that full certifiability is not possible, i.e., agents can disclose some, but not all of their information (see, e.g., Okuno-Fujiwara, Postlewaite, and Suzumura (1990), Shin (1994), and Lipman and Seppi (1995)). The main result of these works is that when interested groups are in competition and have opposed interests, less certifiability is required to achieve the full-information decision.

As matters stand, little research has been done on higher order-information transmission, certification possibilities, and endogenous information structures. Such studies are still new and not very well appreciated, but we think that some interesting results and phenomena might be obtained. In the light of all the preceding analyzes, particularly sections 2, 3, and 4, we know that interactive knowledge has often ambiguous welfare effects. Moreover, slight modifications of some parameters of the problem can drastically change the behaviors. An obvious conclusion of these studies is that common knowledge is bad (respectively good and asymmetric) when the equilibria are bad (respectively good and asymmetric). Arbitrary modeling can often easily lead to desired outcomes. Thus, letting agents voluntarily communicate and modify the information structure in order to create or not some levels of mutual knowledge is crucial. Once we can do this, then we can see if some equilibria and configurations of knowledge are more frequently achieved. It might also allow to see how endogenous knowledge evolution depends on communication practices and certification possibilities. This research is still in a nascent stage.

Conclusion

In interactive situations, the information structure is driven by interactive knowledge. Some languages and some practices of communication could create common knowledge or only lower-levels of interactive knowledge, and could drastically influence strategic behaviors. The information structure also depends on certification possibilities. In both these interdependent configurations of the environment where agents interact, welfare effects are always ambiguous, and this paper has shown that one cannot always support an economy with perfect communication possibilities and with favorable conditions of knowledge sharing, at least in some local and short term interaction problems. However, such conditions help for agreement and are prerequisite for various coordinated decisions.

Some results are clearly very strong and were based on least realistic assumptions and conceptualizations. The weakness of common knowledge theories relies on the limits of game theory, and emphasizes the weakness of economic analysis based on them. However, if we do not use them naively, and if we restrict the analysis to small scale applications, many economic behaviors and phenomena might be studied very precisely and properly. What is more, some more general practices such as advertisement, leadership, monitoring, social and organizational communication could be explained and understood in terms of interactive knowledge and strategic behaviors. Some formalization is also useful to find exactly on which assumptions some results rely on . Even if common knowledge is usually difficult to justify theoretically and empirically, it is very important to analyze it with a view to understanding phenomena like backward induction, coordination, agreement, and more generally conscious and interactive decision making.

The most embarrassing point in the rigorous use of common knowledge seems to be

its underlying epistemic condition of rationality and full awareness. Nevertheless, recent research in bounded rationality and reasoning may be fruitful. Some bounded rationality might be introduced in cognitive rationality or in instrumental rationality. The latter relates to the manner agents determine and compute their action, by reasoning on their given available actions, on their preferences and on their beliefs, i.e., it characterizes the adequacy between means and objectives. Bounded rationality is often introduced through evolutionary modeling or by applying automata theories. The problem is often such models do not consider cognitive rationality (rationality of knowledge).[43] Our view is that weakening cognitive rationality by keeping perfect instrumental rationality (rationality of action) is also very interesting. Many works go in such a direction, for example by considering non-additive beliefs, non-Bayesian learning, by weakening the knowledge axiomatics (e.g., non-partitional information structures) or by taking into account unforeseen contingencies. Such an evolution will be fruitful to understand and rigorously appreciate interactive behaviors which might be qualitatively different than in the standard models, exhibiting different characteristics and mechanisms in economic problems. As a whole, these approaches highlight two crucial and interdependent grounds in interactive decision making: *interactive knowledge* and *strategic information*.

References

ALLEN, F., AND S. MORRIS (1999): "Finance applications of game theory," in *Advances in Business Applications of Game Theory,* forthcoming, ed. by A. Chatterjee, and W. F. Samuelson. Kluwer Academic Press.

AUMANN, R. J. (1974): "Subjectivity and correlation in randomized strategies," *Journal of Mathematical Economics*, 1, 67–96.

——— (1976): "Agreeing to disagree," *The Annals of Statistics*, 4(6), 1236–1239.

——— (1987): "Correlated equilibrium as an expression of Bayesian rationality," *Econometrica*, 55(1), 1–18.

——— (1995): "Backward induction and common knowledge of rationality," *Games and Economic Behavior*, 8, 6–19.

——— (1996): "Reply to Binmore," *Games and Economic Behavior*, 17, 138–146.

——— (1998): "On the centipede game," *Games and Economic Behavior*, 23, 97–105.

——— (1999a): "Interactive epistemology I: Knowledge," *International Journal of Game Theory*, 28(3), 263–300.

——— (1999b): "Interactive epistemology II: Probability," *International Journal of Game Theory*, 28(3), 301–314.

AUMANN, R. J., AND A. BRANDENBURGER (1995): "Epistemic conditions for Nash equilibrium," *Econometrica*, 63(5), 1161–1180.

AXELROD, R. (1984): *The Evolution of Cooperation.* Basic Books.

AYRES, I., AND B. J. NALEBUFF (1998): "Common knowledge as a barrier to negotiation," mimeo.

---

[43]Cognitive rationality is related to the manner agents revise their knowledge and beliefs, and perceive their world, i.e., it characterizes the adequacy between information and beliefs.

BACHARACH, M. (1985): "Some extensions of a claim of Aumann in an axiomatic model of knowledge," *Journal of Economic Theory*, 37, 167–190.

BACHARACH, M., L.-A. GÉRARD-VARET, P. MONGIN, AND H. S. SHIN (eds.) (1997): *Epistemic logic and the theory of games and decisions.* Kluwer Academic Publishers, Boston, MA.

BARWISE, J. (1988): "Three views of common knowledge," in *Proceedings of the Second Conference on Theoretical Aspects of Reasoning about Knowledge (Pacific Grove, CA, 1988)*, pp. 365–379, Los Altos, CA. Morgan Kaufmann.

BASSAN, B., M. SCARSINI, AND S. ZAMIR (1997): ""I don't want to know": Can it be rational?," Discussion Paper 158, Center For Rationality and Interactive Decision Theory, The Hebrew University of Jerusalem.

BATTIGALLI, P., AND G. BONANNO (1999): "Recent results on belief, knowledge and the epistemic foundations of game theory," *Research in Economics*, 53, 149–225.

BERNHEIM, B. D. (1984): "Rationalizable strategic behavior," *Econometrica*, 52(4), 1007–1028.

BINMORE, K. (1992): *Fun and Games.* D.C. Heath and Company.

——— (1996): "A note on backward induction," *Games and Economic Behavior*, 17, 135–137.

BINMORE, K., AND A. BRANDENBURGER (1990): "Common knowledge and game theory," in *Essays on the foundations of game theory*, ed. by K. Binmore, chap. 4, pp. 105–150. Basil Blackwell.

BINMORE, K., AND L. SAMUELSON (1999): "Coordinated action in the electronic mail game," mimeo.

BRANDENBURGER, A. (1992): "Knowledge and equilibrium in games," *Journal of Economic Perspectives*, 6(4), 83–101.

BRANDENBURGER, A., AND E. DEKEL (1987): "Common knowledge with probability 1," *Journal of Mathematical Economics*, 16, 237–245.

——— (1989): "The role of common knowledge assumptions in game theory," in *The Economics of Missing Markets, Information, and Games*, ed. by F. H. Hahn, pp. 46–61. Oxford University Press, New York.

CAVE, J. A. K. (1983): "Learning to agree," *Economics Letters*, 12, 147–152.

CHWE, M. S.-Y. (1998a): "Believe the hype: solving coordination problems with television advertising," mimeo, Department of Economics, University of Chicago.

——— (1998b): "Culture, circles, and commercials: publicity, common knowledge, and social coordination," *Rationality and society*, 10, 47–75.

——— (1999a): *Rational ritual: culture, coordination, and common knowledge.* Princeton University Press, , forthcoming.

——— (1999b): "Structure and strategy in collective action," *American Journal of Sociology*, 105, 128–156.

———— (2000): "Communication and coordination in social networks," *Review of Economic Studies,* forthcoming.

COWAN, R., AND D. FORAY (1997): "The economics of codification and the diffusion of knowledge," *Industrial and corporate change*, 6(3), 595–622.

CRÉMER, J. (1990): "Common knowledge and the coordination of economic activities," in *The firm as a Nexus of Treaties*, ed. by M. Aoki, and O. E. Williamson. European Sage, London.

———— (1993): "Corporate culture and shared knowledge," *Industrial and Corporate Change*, 2(3), 351–386.

DULLECK, U. (1997): "A note on the e-mail game – Bounded rationality and induction," mimeo, Humboldt University, Institute of Economic Theory.

DUPUY, J.-P. (1989): "Convention et common knowledge," *Revue Economique*, 2, 361–400.

FAGIN, R., J. Y. HALPERN, Y. MOSES, AND M. Y. VARDI (1995): *Reasoning about Knowledge.* MIT Press, Cambridge, Massachusetts.

———— (1996): "Common knowledge: now you have it, now you don't," mimeo.

———— (1999): "Common knowledge revisited," *Annals of Pure and Applied Logic*, 96, 89–105.

FAGIN, R., J. Y. HALPERN, AND M. Y. VARDI (1991): "A model-theoretic analysis of knowledge," *Journal of the ACM*, 91(2), 382–428.

FARRELL, J. (1988): "Communication, coordination and Nash equilibrium," *Economics Letters*, 27, 209–214.

FORGES, F. (1986): "An approach to communication equilibria," *Econometrica*, 54(6), 1375–1385.

FOSS, N. J. (1999): "Understanding leadership: A coordination theory," DRUID Working Paper No. 99–3.

GEANAKOPLOS, J. (1994): "Common knowledge," in *Handbook of Game Theory*, ed. by R. J. Aumann, and S. Hart, vol. 2, chap. 40, pp. 1437–1496. Elsevier Science B. V.

GEANAKOPLOS, J., AND H. M. POLEMARCHAKIS (1982): "We can't disagree forever," *Journal of Economic Theory*, 28, 192–200.

GILBOA, I. (1998): "Counter-counterfactuals," *Games and Economic Behavior*, 24, 175–180.

GILLET, E., AND P. GOCHET (1993): "La logique de la connaissance : le problème de l'omniscience logique," *Dialectica*, 47(2–3), 143–171.

GRAY, J. (1978): "Notes on database operating systems," IBM Research Report RJ 2188.

GROSSMAN, S. J. (1981): "The informational role of warranties and private disclosure about product quality," *Journal of Law and Economics*, 24, 461–483.

HALPERN, J. Y. (1995): "Reasoning about knowledge: a survey," in *Handbook of Logic in Artificial Intelligence and Logic Programming*, ed. by D. Gabbay, C. J. Hogger, and J. A. Robinson, vol. 4, pp. 1–34. Oxford University Press.

————— (1998): "Substantive rationality and backward induction," mimeo, Department of Computer Science, Cornell University.

HALPERN, J. Y., AND Y. MOSES (1990): "Knowledge and common knowledge in a distributed environment," *Journal of the ACM*, 37(3), 549–587.

HARSANYI, J. C. (1967–1968): "Games with incomplete information played by Bayesian players. Parts I, II, III," *Management Science*, 14, 159–182, 320–334, 486–502.

HEIFETZ, A. (1996): "Comment on consensus without common knowledge," *Journal of Economic Theory*, 70, 273–277.

HINTIKKA, J. (1969): *Knowledge and belief.* Cornell University Press, Hithaka, NY.

KOESSLER, F. (2000): "Common knowledge and consensus with noisy communication," Working Paper 2000–05, BETA, Université Louis Pasteur, Strasbourg.

KRASUCKI, P. (1996): "Protocols forcing consensus," *Journal of Economic Theory*, 70, 266–272.

KREPS, D. M., P. MILGROM, J. ROBERTS, AND R. WILSON (1982): "Rational cooperation in the finitely repeated prisoners' dilemma," *Journal of Economic Theory*, 27(2), 245–252.

KRIPKE, S. (1963): "A semantical analysis of modal logic I: normal modal propositional calculi," *Zeitschirft für Mathematishe Logik und Grundlagen der Mathematik*, 9, 67–96.

LACAN, J. (1966): "Le temps logique et l'assertion de certitude anticipée," *Ecrits, Paris, Le seuil*, pp. 197–213.

LEWIS, D. (1969): *Convention, a philosophical study.* Harvard University Press, Cambridge, Mass.

LIPMAN, B. L. (1990): "On the strategic advantages of a lack of common knowledge," in *Theoretical Aspects of Reasoning about Knowledge, Proc. of the Third Conf.*, ed. by R. Parikh, pp. 209–224. Morgan Kaufmann, San Francisco.

LIPMAN, B. L., AND D. SEPPI (1995): "Robust inference in communication games with partial provability," *Journal of Economic Theory*, 66, 370–405.

LITTELWOOD, J. E. (1953): *Mathematical Miscellany.* Cambridge Univ. Press, London, 1986, B. Bollobas.

MARENGO, L. (1992): "Structure, competence and learning in an adaptive model of the firm," Papers on Economics and Evolution no. 9203 edited by the European Study Group for Evolutionary Economics, Université catholique de Louvain.

MILGROM, P. (1981a): "An axiomatic characterization of common knowledge," *Econometrica*, 49(1), 219–222.

————— (1981b): "Good news and bad news: representation theorems and applications," *Bell Journal of Economics*, 12, 380–391.

MILGROM, P., AND J. ROBERTS (1982): "Predation, reputation, and entry deterrence," *Journal of Economic Theory*, 27, 280–312.

MILGROM, P., AND N. STOCKEY (1982): "Information, trade and common knowledge," *Journal of Economic Theory*, 26, 17–27.

MODICA, S., AND A. RUSTICHINI (1994): "Awareness and partitional information structures," *Theory and Decision*, 37(1), 107–124.

——— (1999): "Unawareness and partitional information structures," *Games and Economic Behavior*, 27(2), 265–298.

MONDERER, D., AND D. SAMET (1989): "Approximate common knowledge with common belief," *Games and Economic Behavior*, 1, 170–190.

MORRIS, S., A. POSTLEWAITE, AND H. S. SHIN (1995): "Depth of knowledge and the effect of higher order uncertainty," *Economic Theory*, 6, 453–467.

MORRIS, S., AND H. S. SHIN (1997): "Approximate common knowledge and coordination: recent lessons from game theory," *Journal of Logic, Language, and Information*, 6, 171–190.

MYERSON, R. B. (1994): "Communication, correlated equilibria and incentive compatibility," in *Handbook of Game Theory*, ed. by R. J. Aumann, and S. Hart, vol. 2, chap. 24, pp. 827–847. Elsevier Science B. V.

NEEMAN, Z. (1996): "Common beliefs and the existence of speculative trade," *Games and Economic Behavior*, 16, 77–96.

NEYMAN, A. (1991): "The positive value of information," *Games and Economic Behavior*, 3, 350–355.

NISHIHARA, K. (1991): "A note on the equivalence of the two definitions of common knowledge," *Mathematical Social Sciences*, 21, 177–178.

OKUNO-FUJIWARA, A., M. POSTLEWAITE, AND K. SUZUMURA (1990): "Strategic information revelation," *Review of Economic Studies*, 57, 25–47.

PARIKH, R., AND P. KRASUCKI (1990): "Communication, consensus, and knowledge," *Journal of Economic Theory*, 52, 178–189.

PEARCE, D. G. (1984): "Rationalizable strategic behavior and the problem of perfection," *Econometrica*, 52(4), 1029–1050.

PETTIT, P., AND R. SUGDEN (1989): "The backward induction paradox," *Journal of Philosophy*, 4, 1–14.

RABIN, M. (1994): "A model of pre-game communication," *Journal of Economic Theory*, 63, 370–391.

RENY, P. J. (1993): "Common belief and the theory of games with perfect information," *Journal of Economic Theory*, 59(2), 257–274.

RIBEIRO, S., AND D. C. WERLANG (1989): "Common knowledge," in *The New Palgrave: Game Theory*, ed. by J. Eatwell, M. Milgate, and P. Newman, pp. 74–85. W. W. Norton, New York, London.

ROSENTHAL, R. (1981): "Games of perfect information, predatory pricing, and the chain-store paradox," *Journal of Economic Theory*, 25, 92–100.

RUBINSTEIN, A. (1989): "The electronic mail game: strategic behavior under "almost common knowledge"," *American Economic Review*, 79(3), 385–391.

SAMET, D. (1996): "Hypothetical knowledge and games with perfect information," *Games and Economic Behavior*, 17, 230–251.

——— (1997): "Counterfactuals in wonderland," Discussion Paper 134, Center For Rationality and Interactive Decision Theory, The Hebrew University of Jerusalem.

SAVAGE (1954): *The foundations of statistics.* Cambridge Univ. Press, Cambridge.

SCHELLING, T. (1960): *The Strategy of conflict.* Harvard University Press.

SEBENIUS, J. K., AND J. GEANAKOPLOS (1983): "Don't bet on it: contingent agreements with asymmetric information," *Journal of the American Statistical Association*, 78(382), 424–426.

SELTEN, R. (1978): "The chain-store paradox," *Theory and Decision*, 9, 127–159.

SHANNON, C. E. (1948): "A mathematical theory of communication," *Bell System Technical Journal*, 27, 379–423, 623–656.

SHIN, H. S. (1994): "The burden of proof in a game of persuasion," *Journal of Economic Theory*, 64, 253–264.

——— (1996): "How much common belief is necessary for a convention?," *Games and Economic Behavior*, 13, 252–268.

SHIN, H. S., AND T. WILLIAMSON (1994): "Representing the knowledge of Turing machines," *Theory and decision*, 37, 125–146.

STALNAKER, R. C. (1996): "Knowledge, belief and counterfactual reasoning in games," *Economics and Philosophy*, 12, 133–163.

TAN, T., AND S. WERLANG (1988): "The Bayesian foundations of solution concepts of games," *Journal of Economic Theory*, 45(2), 370–391.

VAN DAMME, E. (1995): "Equilibrium selection in team games," mimeo, CentER for Economic Research, Tilburg University.

WALLISER, B. (1991): "Logique épistémique et théorie des jeux," *Revue Économique*, 5, 801–832.