Economics and Philosophy, **30** (2014) 365–395 © Cambridge University Press
doi:10.1017/S0266267114000261

A FRAMEWORK FOR COMMUNITY-BASED SALIENCE: COMMON KNOWLEDGE, COMMON UNDERSTANDING AND COMMUNITY MEMBERSHIP

CYRIL HÉDOIN
University of Reims Champagne-Ardenne, France
cyril.hedoin@univ-reims.fr

This article presents a community-based account of salience as an alternative and a complement to the 'natural salience' approach which is endorsed by almost all game theorists who use this concept. While in the naturalistic approach, salience is understood as an objective and natural property of some entities (events, strategies, outcomes), the community-based account claims that salience is a function of community membership. Building on David Lewis's theory of common knowledge and on some of its recent refined accounts, I suggest that salience acts as a correlating device in a correlated equilibrium. What is constitutive of salience is common understanding, the fact that agents have common knowledge that they share the same modes of reasoning with respect to a well-identified set of events. I argue that the basis for common understanding is community membership. The relevance of this account stems from the fact that it answers the objection that salience is either unnecessary or unable to account for coordination between rational agents.

A primary version has benefited from the comments of two anonymous referees of this journal and also from the remarks of participants of the 'Convention' seminar at the University of Paris-Ouest-Nanterre-La Défense, the 21/10/2011. I also thank François Maniquet for his helpful editorial advice. Errors and omissions are of my full responsibility.

https://doi.org/10.1017/S0266267114000261 Published online by Cambridge University Press

1. INTRODUCTION

Many game theorists think that a concept of salience is required to account for coordination as a part of a theory of social norms and conventions. Indeed, salience is a way to overcome the indeterminacy problem in game theory. Almost all game-theoretic accounts of salience build on David Lewis's theory of conventions (Lewis 1969) and develop a naturalistic conception of salience. It is then understood as an objective and natural property of some entities (events, strategies, outcomes). The purpose of this article is to develop an alternative and complementary conception of salience, while remaining along the lines drawn by Lewis, in particular through his theory of common knowledge. More precisely, I defend a *community-based* approach of salience here, or more simply *community-based salience*. The property of salience is to generate common knowledge expectations among the participants of a strategic interaction. Assuming that salience is a property of events, I suggest that it acts as a correlating device in a correlated equilibrium. What is constitutive of salience is common understanding, the fact that agents have common knowledge that they share the same modes of reasoning with respect to a well-identified set of events. I argue that the basis for common understanding is community membership. The relevance of this account stems from the fact that it answers the objection that salience is either unnecessary or unable to account for coordination between rational agents. As a result, this article defends a non-naturalist and non-individualistic conception of salience. It argues that in many cases we cannot explain how salience serves as a correlating device without acknowledging that agents are members of well-established communities. This approach thus reintroduces in an original way the claim made by some social scientists regarding 'embeddedness': the fact that to explain social behaviour we have to acknowledge that it takes place in or through networks, communities and social structures. However, I do not intend to mean that all cases of salience are instances of community-based salience; but also I claim that not *all* cases of salience refer to natural salience, contrary to what is sometimes implicitly suggested in the game-theoretic literature on the subject.

The rest of this article is made up of seven parts. In the second section, I briefly review the problem of indeterminacy in game theory and explain why salience as defined by David Lewis is relevant to solve this problem. It also reviews Lewis's theory of common knowledge. Section 3 draws the first lines of the community-based account of salience and it connects salience with the concept of correlated equilibrium. Section 4 builds on the latter connection to relate salience with the common priors assumption in game theory. Section 5 offers the formal account of community-based salience. Section 6 briefly discusses the various versions of 'natural salience' and it explains their major limits. I mention in particular the

https://doi.org/10.1017/S0266267114000261 Published online by Cambridge University Press

conundrum underlined by Margaret Gilbert (1992), regarding the fact that it is not possible to entertain together a naturalistic account of salience with the assumption of fully rational individuals. Section 7 explains that the community-based account avoids these difficulties but also acknowledges that natural salience still has some relevance. A final section concludes.

2. INDETERMINACY, SALIENCE AND COMMON KNOWLEDGE IN GAME THEORY

Economists and game theorists continue to struggle with two kinds of problems related to solution concepts in game theory: multiplicity and indeterminacy (Sugden 1991; Hargreaves-Heap and Varoufakis 2004). The problem of multiplicity stems from the fact that many games have several (pure and mixed) Nash equilibria. Multiplicity poses a serious threat to game theory conceived as a predictive and normative tool. The Nash equilibrium refinement research programme (Harsanyi and Selten 1988) has dealt with this problem but it has remained largely unsuccessful. Indeterminacy is a logical consequence of the multiplicity problem: since there is often more than one Nash equilibrium in a game, the outcome of the game becomes impossible to determinate on a priori grounds. However, the problem of indeterminacy expands beyond this limit: the theorist is not the sole person to face indeterminacy; in fact, the tacit assumption of symmetry between the way the modeller describes the game and the way the players frame the game they are playing implies that players also face indeterminacy. This second limit is severe because then it is simply impossible for the theorist to explain how and why players behave as they do in a specific game. In fact, when there is indeterminacy, players themselves must be unable to choose. Indeterminacy goes even further once we realize that the traditional assumption of common knowledge of rationality among players does not entail that one of the Nash equilibria will be played.

As early as 1960, Thomas Schelling produced empirical evidence that most of the time people are able to coordinate not only in coordination games but also in 'mixed-motive games' where the interests of the players are not perfectly aligned. Schelling's aim was to show that in games where the mathematical structure seems to leave players unable to make any choice but a random one, individuals demonstrate an ability to coordinate on some specific outcomes. According to Schelling, people are able to 'solve' coordination and mixed-motive games because they rely on *focal points*. Focal points are defined by their 'conspicuousness' or by their prominence. A few years later, the philosopher and logician David Lewis built on Schelling's insights to develop his theory of social conventions (Lewis 1969). In the process, Lewis also developed one of the very first accounts of *common knowledge*. The way Lewis articulates his theory of

https://doi.org/10.1017/S0266267114000261 Published online by Cambridge University Press

common knowledge with the concept of prominence or, as he calls it, 'salience', is highly interesting but it has not attracted much attention until recently. As I wish to make clear, Lewis's theory of common knowledge is of the utmost importance because it can help us to gain a better understanding of the epistemic foundations of salience. In turn, these epistemic foundations will allow us to go beyond the idea of salience as a mere natural phenomenon. The starting point of Lewis's study is the claim made by the philosopher W.V. Quine that language cannot be a convention since conventions need language to be established. Lewis's aim was to demonstrate that conventions may be established without any kind of communication device, and therefore that language could be conventional.

Lewis defines conventions as devices solving coordination problems. A convention is a 'coordination equilibrium' in a game where there are at least two coordination equilibria.[1] Now, the following question arises: since a convention is an equilibrium in a (coordination) game with at least two equilibria, how can any member of a population P be sure of the regularity which will be followed by other members of P? This is the very same problem of coordination studied by Schelling and in fact Lewis produces a similar solution. Lewis sees that the resolution of all coordination problems involves the generation of *higher-order expectations* by individuals. It means that, starting from the knowledge of individuals' rationality and preferences, each person in a coordination problem will draw expectations on what others expect on what everyone expects, and so on. More formally, define A as a bundle of information on the rationality, preferences and action of each member of P; a first-order expectation 'I expect A' is produced by higher-order expectations of the kind 'I expect that you expect A', 'I expect that you expect that I expect A', 'I expect that you expect that I expect that you expect A', and so on. If we note $R^n(A)$ as for 'I expect A' with n the n-order of expectations, then the decision to conform to A as the result of the first-order expectation $R(A)$ is produced by an infinite chain of higher-order expectations with n tending to infinity.[2] Of course, as Lewis notes, human beings hardly have the capacity to generate such an infinite chain of expectations. More often than

[1] Lewis defines a coordination equilibrium 'as a combination in which no one would have been better off had *any one* agent alone acted otherwise, either himself or someone else' (1969: 14). This definition is restrictive and applies only to coordination games. More recent studies of conventions drawing on Lewis's work adopt a more liberal definition. For example, Robert Sugden (2004) defines a convention as any equilibrium in a game with at least two (Nash) equilibria.

[2] For a reconstruction, see, for example, Sillari (2005) or Rescorla (2011). There have been several objections against Lewis's account of common knowledge. See in particular the criticism developed by Margaret Gilbert (1992). It is no coincidence that Gilbert also rejects Lewis's claim that salience allows individuals to solve coordination problems. I return to this point later.

https://doi.org/10.1017/S0266267114000261 Published online by Cambridge University Press

not, expectations beyond a certain order are unnecessary, but the higher order the expectations are, the more one has a reason to believe something and to act in a certain way (Lewis 1969: 33). One of the main achievements of Lewis's study is to demonstrate how such a chain of expectations might be generated to make coordination possible.

Lewis first proposed a theory of common knowledge. The aim is to determine necessary and sufficient conditions for a state of affairs A to make a proposition common knowledge in a population. Lewis informally states those conditions as follows (1969: 52–53): provided that everybody shares the same background information and the same inductive standards, a state of affairs A can make a proposition P common knowledge if (1) each of us has reasons to believe that A holds, (2) A indicates to everyone that each of us has reasons to believe that A holds, (3) A indicates to everyone that P. The proposition 'A indicates to everyone that P' means that if I have reasons to believe that A holds, then I have reason to believe that P is true. A and P are not necessarily linked deductively; the link may rely on some inductive standard and on some background information. Lewis demonstrates that these premises, together with reasonable premises on the rationality, inductive standards and background information of the agents, are sufficient to generate the infinite chain of higher-order expectations. It is useful to formalize the epistemic conditions that are necessary for the higher-order expectations to be generated. For all persons i, j in a given population, a proposition P is common knowledge at a state of the world $\omega \in \Omega$, if and only if for a state of affairs A such that $\omega \in A$, the following four conditions are satisfied:[3]

(1) $\omega \in K_i(A)$
(2) $K_i(A) \subseteq K_i[K_j(A)]$

[3] K is a knowledge operator. I assume in this article that the knowledge operator satisfies the S-5 axioms of modal logic as is traditionally the case in epistemic logic or as in the set-theoretical treatment of common knowledge by Aumann (1976). Gintis (2009) and Vanderschraaf (1998) also both use standard modal logic to formalize Lewis's treatment of common knowledge. Cubitt and Sugden (2003) have criticized the use of the set-theoretical approach to formalize Lewis's theory. Among the various arguments they develop, they argue that Lewis's theory is not about knowledge but about belief (more precisely, about *reason to believe*). More importantly, they claim that the set-theoretical approach leads to reduce Lewis's notion of relation of indication to a mere material implication (i.e. set inclusion in set-theoretical terms) and as a consequence to ignore the possibility that agents may entertain different modes of reasoning. Sillari (2005) and Paternotte (2011) have proposed two middle-way approaches, keeping with the modal logic framework while incorporating a proper formalization for Lewis's two key notions of indication and reason to believe. Since this paper is primarily about salience, I do not intend to take position here regarding what is the most relevant formalization of Lewis's theory of common knowledge. As far as I can tell, the 'simpler' (most conventional) set-theoretical approach is sufficient to make my own point. Thus, I keep this approach here but the reader should have in mind the above subtleties.

https://doi.org/10.1017/S0266267114000261 Published online by Cambridge University Press

(3) $K_i(A) \Rightarrow K_i(P)$
(4) $K_i(A) \wedge K_i[K_j(A)] \Rightarrow K_i[K_j(P)]$

These four conditions read as follows: the first condition states that each person i knows (has reason to believe) that A holds in the current state of the world; condition 2 indicates that if i knows that A holds, then he knows that j knows that A holds. Condition 3 is the formal statement of Lewis's relation of indication: under a particular mode of reasoning (possibly but not necessarily deduction), the agent i infers P from A. The fourth condition states that a person i can infer from his knowledge of the public character of A that j knows P. This condition makes explicit a crucial assumption in Lewis's theory: individuals are *A-symmetric reasoners*, i.e. they use the same kind of mode of reasoning and they know this fact with respect to A (Vanderschraaf and Sillari 2009). The first two conditions are purely empirical: the fact that A happened may or may not lead me to know that it happened and to know that you know that it happened. Together, they imply that A is a *public* event. Conditions 3 and 4 are epistemic: clearly, they necessitate that everyone uses similar reasoning modes and that everyone is sufficiently confident of this fact. If these four conditions are satisfied, then the following general proposition is true for the set of agents N:

(5) $\omega \subseteq K_N^*(P)$

Proposition (5) says that at ω, an event A makes a proposition P common knowledge in a given population constituted by a set of agents N.[4] (5) defines an iterative chain of 'i knows that j knows that … knows that P'. An equivalent but more compact way to express common knowledge is to use a fixed point formulation (Vanderschraaf and Sillari 2009) where all agents in N know that P and know that all know that P (i.e. for a set of agents N, $K_N[P \wedge K_N(P)]$).

When (5) is true, A is said to be an *indicator* or a *reflexive common indicator* of P in N, or equivalently, A is the *basis* for the agents in N of the common knowledge of P. Lewis uses the salience of an event or state of affairs A as an explanation of the fact that A rather than any other event A' will make P common knowledge.[5] More exactly, what makes an event or a state of affairs A salient is the fact that it is an indicator

[4] See Gintis (2009: 142, theorem 7.2). Obviously, given conditions 2 to 4, for two agents i, $j \in N$, we have $K_i(A) \subseteq K_i[K_j(P)]$. Since $\omega \in K_i(A)$, we have $\omega \in K_i[K_i(P)]$. By the same logic, for any person i, j, $k \in N$ and given conditions 2–4, we have $K_i(A) \in K_i[K_j[K_k(P)]]$, which gives $\omega \in K_i[K_j[K_k(P)]]$. The same is true for any level of mutual knowledge in N, thus proving common knowledge.

[5] One of the referees remarked that it is not at all clear that Lewis attributed salience to events rather than to equilibria. Indeed, when discussing salience, Lewis (1969: 35–42)

https://doi.org/10.1017/S0266267114000261 Published online by Cambridge University Press

of P in N. Indeterminacy is solved when agents are able to form a set of hierarchical expectations that converge toward an equilibrium. Such a convergence is made possible if A is a reflexive common indicator. Similarly, as an explicit agreement, the salience of an event makes people know (or believe) that everyone will behave according to what this event indicates. Hence, the salience of an event is defined by its property to make a proposition or a set of expectations common knowledge. Salience is important in Lewis's theory because it helps to account both for the origins of conventions and for their perpetuation through time (Sillari 2008). Regarding the maintenance of the convention, Lewis essentially puts emphasis on the force of the precedent as a special kind of salience allowing one to expect that everyone will follow a regularity R from the fact that regularity held in the past: given a reasonable inductive standard, the state of affairs A that R has been followed gives everyone a reason to believe P that R will be followed the next time. However, as Lewis himself recognizes, the same reasoning is valuable for any kind of salience.

The link between salience and symmetric reasoning in Lewis's account is tacit but essential. As condition 4 formally states, for a proposition to become common knowledge the members of a population must be symmetric reasoners. Also, though this is implicit, they must know the latter. This means that a salient event must be an event for which it is obvious (no matter how this obviousness might be explained) that it will indicate the same proposition to everyone.[6] Lewis takes symmetric reasoning as a reasonable assumption but does not give the necessary or sufficient conditions for it to be verified. If we follow Lewis's path, then any explanation of salience should ask how and when symmetric reasoning can be obtained. Lewis seems to consider symmetric reasoning essentially as the sharing of obvious or 'natural' forms of reasoning. For instance, the obviousness of a convention lies in the fact that it is a recurring practice and that it seems *natural* to expect that what has worked in the past should work in the future. Clearly, the precedent

tends to speak of the salience of a coordination equilibrium. However, I tend to think that Lewis entertains some confusion on this point because we still have to distinguish between the prevailing state of affairs (e.g. this convention has been followed in the past) and the conclusions or expectations that the agents may form on its basis (e.g. everyone will continue to follow the convention). But, at least in Lewis's discussion of salience because of precedent, it should be acknowledged that the salience of the convention (the equilibrium) proper follows from the salience of the state of affairs (the precedent). In the remainder of this paper, I will thus assume that salience is a property of both the event and of the resulting equilibrium through this kind of connection.

[6] If individuals are not symmetric reasoners, then an event A cannot make a proposition P common knowledge. Therefore, A cannot be salient because nobody will have reason to expect that others will know P. It is obvious that the condition of symmetric reasoning is needed to make salience a way to overcome indeterminacy. It is therefore crucial to give a convincing account of symmetric reasoning.

https://doi.org/10.1017/S0266267114000261 Published online by Cambridge University Press

makes this solution salient because individuals are deemed to share a common inductive form of reasoning where past events are taken as a prediction of what future events will be. Hence, this 'intuitive' form of reasoning gives a particular salience to the precedent. It appears that salience because of precedent is a particular case of *natural salience*. In the rest of the paper, I will defend another (but not incompatible) conception of salience, *community-based salience*. I interpret symmetric reasoning as a matter of common understanding where individuals share modes of reasoning and know this fact. I argue that the condition of symmetric reasoning is satisfied in many cases because individuals pertain to the same communities and interact through the same institutions.

3. CORRELATED EQUILIBRIUM AND (COMMUNITY-BASED) SALIENCE

Because Lewis tries to explain how a given fact or proposition x becomes commonly known in a population, he provides the epistemic foundations for a complete understanding of salience. But Lewis's theory remains incomplete because it does not explain *why* a specific event A became an indicator while a similar but not identical event A' fails to generate common knowledge of expectations in the same population. Indeed, salience is still an *explanans* and Lewis did not attempt to provide an *explanation* of it. Nevertheless, Lewis's theory of common knowledge gives a critical insight by indicating that salience operates through the fact that individuals are symmetric reasoners. In this section, I use the concepts of correlated equilibrium and correlated strategies in order to argue that a correlated device will be used because of its salience. Another motivation for the use of the correlated equilibrium concept is that it points to an interesting parallel between salience, symmetric reasoning and the much discussed assumption of common priors. I will return on this last point in the next section.

In the preceding section, I established that an event will be salient if (i) it is public or can be expected to be so (that is, everyone can confidently expect that everyone else has observed the event) and (ii) everyone infers the same conclusion from it (and knows this fact). In a game-theoretic framework, it is generally assumed that the players' beliefs regarding what others will do in a strategic interaction are probabilistically independent. However, the property of salience is precisely to make players' beliefs probabilistically *dependent* upon an event. It is then useful to think of salience as a particular kind of correlating device since it operates by correlating each player's conjectures regarding what others will do. Furthermore, this will help to make clearer our latter point that salience cannot always be reduced to individual (or 'natural') properties. Hence, salience can be understood through the concept of *correlated*

https://doi.org/10.1017/S0266267114000261 Published online by Cambridge University Press

	Sell	Do not sell
Sell	5	6
Do not sell	0	10

FIGURE 1. The Market Panic Game

equilibrium.[7] A *correlating device* is an object that sends out signals (either public or private) to agents indicating which (pure) strategy they should play given the state of the world. Informally, it is analogous to an external agent observing Nature's choice for the state of the world and then sending an instruction to each player of the game on how to play. There is a correlated equilibrium if, for each possible state of the world, no player has incentive to deviate from the instruction of the correlating device when everybody is following it. Formally, a correlated equilibrium is a vector of strategies that is a function of the state of the world and such that, given players' prior beliefs over the possible states of the world and over others' strategies, everyone is maximizing expected utility. At equilibrium, players' beliefs and strategies are common knowledge and are correlated, i.e. are not formed independently.

I will illustrate the idea by a simple example that will also be useful to introduce the idea of 'community-based' salience that I defend in the remainder of the paper. In a recent paper reviewing the contribution of neuroeconomics and of behavioural economics, David Levine (2011) argues that rational behaviour can cause financial bubbles. He therefore estimates that we do not need to look at people's psychology and brain functioning since a theory of irrational behaviour is unnecessary. He gives a sensible illustration (298–299):

> The situation in a market panic is similar. Suppose you turn on the television and notice the Chairman of the Federal Reserve Board giving a speech indicating that the financial sector is close to meltdown. It occurs to you that when this happens, stocks will not have much value. Naturally you wish to sell your stocks – and to do so before they fall in price, which is to say, to sell before everyone else can rush to sell. So there is 'panic' as everyone rushes to sell. Individual behavior here is rational.

The market panic situation described by Levine has the form of an assurance game (see Figure 1; payoffs are those of the row player, the columns represent all the other players).

In this situation, everyone will be better off by not selling his stocks but if others decide to sell, the best one can do is to sell. That everyone rationally chooses to sell after having seen on television the Chairman

[7] The connection between salience and correlated equilibrium has already been drawn several times. See in particular Gintis (2009) and Vanderschraaf (1998).

https://doi.org/10.1017/S0266267114000261 Published online by Cambridge University Press

of the Federal Reserve predicting a financial meltdown is an instance of correlated equilibrium: the speech of the Chairman on television is a (public) signal allowing correlated strategies. It indicates to each player to sell his stocks and clearly, each player has to follow the signal provided that everyone else does the same: the correlated strategy 'sell if the Chairman announces a financial meltdown' is a best reply to itself and so corresponds to a correlated equilibrium. The formalism of Lewis's theory of common knowledge is useful here. The Chairman's speech is a public event A: the fact that the Chairman gives his speech on television leads everyone to reasonably believe that everyone knows A. The public announcement of the Chairman leads each individual to believe the proposition P that a financial meltdown is coming, or more exactly that everyone will sell his stocks. Provided that each individual is an A-symmetric reasoner, everyone knows that everyone knows P: that a financial meltdown is coming becomes common knowledge. All four conditions for common knowledge to be generated are satisfied. Then, the Chairman's speech A is a reflexive common indicator making everyone's expectations converge to P.

It is interesting to ask what makes the Chairman's speech a basis for the common knowledge that everyone will sell. To answer this question is important because it is precisely this property that makes the Chairman's speech a salient event. Indeed, there are potentially an infinite number of signals that people could daily perceive regarding the dynamics of financial markets. What makes the Chairman's speech prominent among the many signals that are sent through the newspapers and television? At the technical level, the answer has already been given in the preceding section: people are A-symmetric reasoners, i.e. people are symmetric reasoners with respect to the Chairman's speech constituting the event A. They share the same modes of reasoning and have reason to believe this fact. Clearly, this cannot be true for all the other signals making a prediction on the evolution of asset prices. If I hear a journalist, I can hardly infer something about what others will infer.[8] Different people may have different opinions about the truth-value of the journalist's announcement, eventually as a function, for example, of the channel he is working for. And different people may have different opinions regarding the overall reputation of the journalist in the population. While I may believe that the journalist is right, I may also entertain the belief that given

[8] I assume here that there is not any information asymmetry between agents and in particular that everyone has the same information regarding the expertise of the journalist. This is an important assumption because differences in information are generally the favoured argument of economists to explain disagreement between agents. I return on this point in the next section since it is related to the whole discussion about the common priors assumption.

https://doi.org/10.1017/S0266267114000261 Published online by Cambridge University Press

he has made wrong predictions in the past, others in the population will believe that what will actually happen is exactly the contrary of what the journalist is predicting. In any case, assuming symmetric reasoning with respect to the journalist's prediction is difficult and therefore this prediction cannot be the basis for common knowledge in the population.

I think it is clear that *institutional* factors are at stake here to explain the epistemic properties of the Chairman's speech. It is clear in this example that we cannot understand how the event A makes the proposition P common knowledge if we ignore the institutional context that is tacitly assumed. In fact, this example is full of institutional signs that make the example understandable and meaningful for everyone. For instance, one cannot understand how the Chairman's speech becomes an indicator if one ignores the institutional role of the Federal Reserve and of its Chairman in the US economy. Now, it is doubtful that the property of A to make P common knowledge is tied to the personality of the Chairman. It is more likely that this property is best explained by the *institutional role* that is occupied by that person. That the head of the Federal Reserve is an important person susceptible to have influence on the economy and who has access to important information is the product of the institutional structure of the US economy. That every American adult person knows this information can be explained by the fact that this is part of a common culture which is itself generated by the fact that Americans live in the midst of the same economic and non-economic institutions. In a nutshell, people are A-symmetric reasoners in this case because they are members of the same community, defined by a set of shared institutions, and that it is in some way or another taken for granted that all the members of the community will infer the same conclusion from the Chairman's speech. I next develop the concepts of common understanding and shared reflexivity to generalize this account of community-based salience. However, I will first briefly argue in the next section for the connection between salience and the much debated assumption of common priors in game theory.

4. A SHORT DIGRESSION: WHERE DO COMMON PRIORS COME FROM?

The second motivation for associating salience with the concept of correlated equilibrium is that this approach makes obvious the connections between salience and symmetric reasoning on the one hand, and the so-called assumption of 'common priors' on the other hand. The common prior assumption holds that rational agents draw the same inferences on how a game is to be played. Meanwhile, the salience of an event makes everyone's expectations regarding how the game will be played common knowledge. Hence, assuming that a strategy

https://doi.org/10.1017/S0266267114000261 Published online by Cambridge University Press

is played because of what a salient event or state of affairs indicates amounts to assuming that players have common priors on the way the game will be played. The association between the common priors assumption and the concept of correlated equilibrium is due to an important theorem, discovered by Robert Aumann (1987), that establishes that Bayesian rational players with common priors in any epistemic game G will implement a correlated equilibrium in the corresponding game G^* augmented by an initial move made by Nature (see also Gintis 2009: 136–139).

An epistemic game G is a normal form game with a set $N = (1, ..., n)$ of n players, a set $S = (S_1 \times ... \times S_n)$ of pure strategies and a function $u_i : S- > R$ mapping any combination of strategies into a n-tuple of real numbers. In addition, G includes a set Ω of possible states of the world, a knowledge partition P_i of Ω for each player and finally a subjective prior $p_i(.;\omega)$ that defines each player's beliefs regarding how the game will be played for each state of the world $\omega \in \Omega$ (Gintis 2009: 83–84). A player's subjective prior states his beliefs concerning the actual state of the game when the actual state of the world is ω. Each state ω specifies the strategy profile that will be used in the game. Hence, a player's subjective prior defines his beliefs over the strategy profile that will be played by others. For a given state of the world ω, a Bayesian rational player will choose the strategy that maximizes his expected payoff given his conjectures on how the other players will play. To say that players have common priors over Ω is simply to say that they have the same beliefs distribution over each possible state of the world ω. Aumann's theorem then demonstrates that when Bayesian rational players have common priors over Ω, they will implement a strategy profile that corresponds to a correlated equilibrium. The converse is also true: any correlated equilibrium in a game G^* corresponds to an epistemic game G where players have common priors such that for each state of the world ω, it is rational for them to play their part in the correlated equilibrium (Gintis 2009: 136). Regarding the working of salience, this theorem is of great significance. We have suggested above that salience works as a correlating device: an exogenous feature helps the players to make convergent and consistent expectations regarding what everyone will do. Salience allows players to form their expectations in a non-independent way. Aumann's theorem demonstrates that it is the same as to assume that (Bayesian rational) players have common priors over Ω, even when there is no explicit correlating device. Hence, the problem of the origins of salience can be at least partially resolved in determining where common priors come from.

The assumption of common priors solves the problem of indeterminacy that occurs when two or more equilibria exist in a game. In fact, assuming common priors amounts to assuming that players

https://doi.org/10.1017/S0266267114000261 Published online by Cambridge University Press

have already agreed on the way they will play. If this agreement corresponds to an equilibrium, then nobody has interest to deviate, provided that others implement the agreement. The common priors assumption has been much debated, in particular in the context of game theory where probabilities about states of the world are endogenous (Morris 1995). Since Leonard Savage's *The Foundations of Statistics* (1954), it is traditional to distinguish between three accounts of probabilities: the logical account, the frequentist account and the subjective account. It is generally considered that assuming common priors for logical and frequentist probabilities is unproblematic. In both cases, probabilities are considered to be an objective property of the world whose knowledge is accessible to any minimally rational agent. For instance, probabilities in casino games or in games such as poker or backgammon can be understood either as logical or frequentist probabilities. In both accounts, they are objective data about the world. Matters are different in strategic interactions because probabilities are about agent's beliefs and actions. But of course, as I have already emphasized above, each agent's beliefs about others' actions depend on other's beliefs regarding his own beliefs. In other words, beliefs and thus probabilities are generated endogenously and assuming logical or frequentist probabilities is simply nonsense. It is well known that Savage developed the subjective account of probabilities to deal with one-person decision problems with uncertainty (i.e. cases where it is impossible to define objective probabilities). However, it has been argued (e.g. Sugden 1991; Hollis and Sugden 1993) that the method proposed by Savage to infer subjective probabilities from the agent's choices and preferences is not valid in the case of strategic interactions for very foundational reasons. Therefore, it is no surprise that the common prior assumption in game theory has been disputed. On the other hand, its rejection seriously undermines the relevance of some solution concepts in game theory, in particular the Nash equilibrium solution concept.

Interestingly, it seems that the common priors assumption leads unavoidably to a departure from any strict individualistic understanding of rationality in a game because it seems difficult to figure how multiple individuals can entertain the same subjective priors in strategic interactions without any kind of intersubjective relationship. However, economists have originally tried to justify this assumption without making any appeal to some form of 'collective' rationality. For instance, John Harsanyi has argued that two rational individuals who are submitted to the same information must necessarily and independently come to the same conclusion.[9] Robert Aumann (1976), who dubbed this claim the

[9] As Stephen Morris (1995) documents, game theorists have developed several arguments to justify the use of the common priors assumption. The most convincing build on methodological and pragmatic justifications. In particular, it is often claimed that allowing

https://doi.org/10.1017/S0266267114000261 Published online by Cambridge University Press

'Harsanyi doctrine', noted that this is the same as to assume that rational individuals have common priors. The Harsanyi doctrine implies that any disagreement between two rational individuals originates necessarily in the differences in the information they rely on. Pushing Harsanyi's claim further, Aumann (1976) demonstrated that if rational individuals have common priors and if their posterior beliefs are common knowledge, then these players cannot 'agree to disagree': because of the Harsanyi doctrine, any disagreement must result from the fact that at least one individual has private information; since the disagreement makes common knowledge individuals' beliefs, it reveals that at least one of them possesses some private information. It follows that if individuals are Bayesian rational, they must entertain the same beliefs.

This last theorem *assumes* that individuals have common priors. The fact that disagreements over many topics in modern societies are pervasive seems to indicate that either individuals do not have common priors or that their posterior beliefs fail to be common knowledge. Indeed, it seems that the Harsanyi doctrine applies only in a narrow set of circumstances. We can only expect individuals to have common priors for straightforward and recurrent problems which qualify for the use of logical or frequentist probabilities. In particular, people probably have common priors regarding several *natural events* such as, for instance, the probability that a coin falls on head. People probably also entertain similar beliefs regarding the probability that a person jumping from the top of a building survives. These types of events involve basic physics laws or obvious consequences such that it is reasonable to expect that all persons endowed with normal mental faculties will hold similar beliefs. Matters are less clear for very rare natural events and for many situations where individuals have to form beliefs over others' intentions or expectations. Here, it seems heroic to assume common priors along the line of the Harsanyi doctrine, unless we can establish that, in one way or another, individuals have been able to tacitly agree over consistent and convergent beliefs.

The similarity between the Harsanyi doctrine and the assumption of symmetric reasoning (section 2) is striking. Indeed, the Harsanyi doctrine states that two rational individuals reason in an identical way. In other words, it amounts to assume that everyone infers the same conclusion from the same set of premises. As a consequence, we can establish the

heterogeneous priors is predisposed to lead to an 'anything goes' type of modelling where any phenomenon can be accounted for by setting the 'right' priors. Note however that such pragmatic justifications neither provide nor try to provide an explanation for the fact that people would have common priors. Harsanyi's argument is of a different nature because it makes of common priors a property of the agents' rationality.

https://doi.org/10.1017/S0266267114000261 Published online by Cambridge University Press

following proposition on which I build in the rest of the text (a short proof is given in the appendix at the end of the text):

> Proposition 1: The common prior assumption and the assumption of symmetric reasoning are formally equivalent.

The Harsanyi doctrine and the assumption of symmetric reasoning pose no problem regarding a small set of natural or basic events. The 'natural' salience of some phenomena falls in that category: for example, people have the ability to uncover some natural patterns in a list of numbers (and everyone can expect everyone to have this ability) because these patterns follow both from a basic knowledge of arithmetic and from basic human mental capacities. As we noted in section 2, symmetric reasoning in many social interactions calls for an explanation. Similarly, the assumption of common priors on which the Harsanyi doctrine is grounded is all but straightforward in complex social interactions. Therefore, it seems that providing an explanation for symmetric reasoning can also provide a partial justification for the assumption of common priors different from the pragmatic justifications usually entertained.

5. COMMUNITY-BASED SALIENCE: A CONCEPTUAL FRAMEWORK

I suggested that the property of a salient event is to generate a set of common knowledge expectations among the members of a population. Following Lewis's theory of common knowledge, this ability is grounded on the assumption of A-symmetric reasoning, i.e. the fact that members of the population are symmetric reasoners with the respect to the salient event A. Alternatively, as I have shown in the previous section, salience can be thought as a peculiar instance where the common priors assumption holds: the event A brings about a new information from which everyone infers the same conclusion, and everyone knows this. I now want to give a more explicit formulation of this assumption and to explain how it is related to the community-based approach of salience, in contrast with the natural salience approach.

In Lewis's account, what is required is not only that individuals share modes of reasoning with respect to a given state of affairs; but also that they have a reason to believe this fact. This is expressed in conditions 3 and 4 in Lewis's account of common knowledge. For any two individuals i, j, for an event or a state of affairs A and for a proposition P, the *Lewis's A-symmetric reasoning condition* states that from

$$\text{A) } K_i(A) \Rightarrow K_i(P)$$

and

$$\text{B) } K_i(A) \Rightarrow K_i[K_j(A)]$$

https://doi.org/10.1017/S0266267114000261 Published online by Cambridge University Press

we must have

$$\text{C) } K_i[K_j(P)]$$

Obviously, there are two tacit epistemic assumptions in the Lewis's *A*-symmetric reasoning condition. The first one is that both *i* and *j* must infer the same conclusion from the same events or states of affairs. Generalizing for a population constituted by the set of agents *N* and denoting K_N for 'everyone in *N* knows that …', we have

$$\text{a) } K_i(A) \Rightarrow K_i(P) \leftrightarrow K_N(A) \Rightarrow K_N(P)$$

It is must be clear that assumption (a) is reminiscent of the Harsanyi doctrine that if two individuals have the same information regarding a state of affairs *A*, then they must reach the same conclusion *P*. But this is not sufficient because for *i* to infer that *j* knows *P* from his own knowledge that *A* indicates *P*, he must also know (or have reason to believe) that *j* used the same mode of reasoning. Or, in other words, everyone in *N* must know assumption (a), i.e.

$$\text{b) } K_N[K_i(A) \Rightarrow K_i(P) \leftrightarrow K_N(A) \Rightarrow K_N(P)]$$

Assumption (b) indicates that each member *i* of *N* knows that if he infers *P* from *A*, then everyone else in *N* also infers *P* from *A*. Note that (b) simplifies into $K_N[K_N(A) \Rightarrow K_N(P)]$ which is to say that sharing reasoning modes is mutual knowledge in *N*. Moreover, note that for producing an additional level of mutual knowledge of *P*, i.e. *i* knows that *j* knows that *i* knows *P*, assumption (b) is not sufficient. Indeed, *i* must also know that *j* knows assumption (a). The same is true for any additional level of mutual knowledge, which means that for a common knowledge of *P* among *i* and *j* to be obtained they must also have common knowledge of the fact that they share the same modes of reasoning.[10] By this reasoning, for *P* to become common knowledge in *N*, we need an assumption (b*) where assumptions *a* and *b* are mutual knowledge in *N*. Assumption (b*) defines a hierarchy of propositions of the kind '*i* knows that *j* knows that *k* knows that … infers *P* from *A*' for any *i*, *j*, *k*, … ∈ *N*. Alternatively, we can use a fixed-point formulation of the common knowledge of reasoning modes in *N*:

> b*) $K_N^*(Q) \leftrightarrow K_N[Q \wedge K_N(Q)]$ with proposition *Q* = everyone shares the same reasoning modes in *N*

I propose to name assumption (a) and assumption (b*) respectively with the concepts of *shared (or common) reflexivity* and of *common understanding*.

[10] As noted by Sillari (2008: 31), 'in [Lewis's] definition of common knowledge itself there seems to be a problematic requirement of some sort of pre-existing common knowledge'.

https://doi.org/10.1017/S0266267114000261 Published online by Cambridge University Press

The notion of 'reflexivity' means that individuals facing a decision problem are able to reflect on their mode of reasoning.[11] By shared reflexivity, I mean the following

> (Φ - N)-*Shared Reflexivity*: there is (Φ - N)-shared reflexivity among the members of a population constituted by the set of agents N if the agents share the same reasoning modes with respect to a non-empty set of events Φ. In practice, all agents in N infer the same conclusion P when they observe an event $A \subseteq \Phi$.

From this definition, it is possible to define a state of n-th order of (Φ - N)-shared reflexivity as the state where each member of N knows that there are n-1 orders of (Φ - N)-shared reflexivity. With respect to assumption (b*), I now define common understanding as the limit case where shared reflexivity is common knowledge:

> (Φ - N)-*Common Understanding*: there is (Φ - N)-common understanding among the members of a population constituted by the set of agents N if it is common knowledge among the members of N that they share the same reasoning modes with respect to a non-empty set of events Φ. In practice, all agents in N infer the same conclusion P when they observe an event $A \subseteq \Phi$ and this is common knowledge.

Given what has been said in section 2, the set of events Φ has a straightforward interpretation. Indeed, Φ contains all the *salient events* in a population N. In other words, all events in Φ have the property of being common reflexive indicators and thus of generating common knowledge expectations. Alternatively, one might describe the set Φ as containing all correlating devices where agents' subjective priors are such that they are able to coordinate on a correlated equilibrium. Thus, (Φ - N)-*Common Understanding*, or more simply common understanding, explains why some correlating devices are salient such that they generate correlated equilibria, while others remain unnoticed. For instance, in Levine's example of the Chairman's speech, the interpretation of the prediction of a financial meltdown on television is obvious to anyone who is acquainted with market economies that have a long history of financial turbulences. Arguably, all living people in market economies have a common understanding of the way these economies work and so understand that a signal sent by the Chairman of the Federal Reserve is significant and meaningful to everyone.

[11] Reflexivity is also a technical assumption in almost all systems of modal logic. Both S-4 and S-5 axioms of modal logic contain an axiom of reflexivity, i.e. $K_i(A) \subseteq K_i[K_i(A)]$. The difference here is that we are not concerned with reflexivity about knowledge but reflexivity about reasoning.

https://doi.org/10.1017/S0266267114000261 Published online by Cambridge University Press

A few qualifications regarding the concept of shared reflexivity and common understanding are to be made before advancing further. In particular, what is true about the status of 'knowledge' or 'reason to believe' in Lewis's account of common knowledge remains true here. Lewis often emphasizes that reasons to believe are not actual beliefs. The hierarchy of mutual knowledge or reasons to believe corresponds to a chain of logical implications, not to steps in one's actual reasoning.[12] The point is that, in practice, not only are people generally unable to reason through the multiple levels of order of mutual knowledge but moreover it is more often than not unnecessary. For instance, to drive on the right side of the road, it will be sufficient for me in practice to know that you will drive on the right side and possibly to know that you know that. Any higher order of mutual knowledge is not really required to give me practical reasons to drive on the right side. Since for any n-th order of mutual knowledge of a proposition P, a parallel n-th order of mutual knowledge of the sharing of reasoning modes is needed, the same conclusion is true regarding shared reflexivity and common understanding. In practice, common understanding needs only to obtain regarding reasons to believe but it is not always required as an *actual* state.[13]

A second issue relates to the origins of common understanding. For a complete explanation of salience and focal points, questions such as 'Where does common understanding come from?' or 'How are higher-orders of shared reflexivity generated?' need to be answered. As suggested in the third section with the example of the market panic

[12] See Sillari (2005) who uses the concept of 'awareness structures' to distinguish reasons to believe from actual beliefs. Basically, the idea is that an agent actually believes a proposition x if he has reason to believe x and if he is aware (in a particular sense) of x. In Sillari's framework, following Lewis's account, the degree of awareness of an agent depends on his level of rationality.

[13] Clearly, because of the symmetry between Lewis's common knowledge and common understanding, one might oppose to both concepts that neither one is actually needed to account for the observed behaviour of people. Binmore (2008), among others, argues that Lewis's common knowledge is unnecessary to explain conventions. However, this is not sufficient as an argument to reject Lewis's common knowledge and my related concept of common understanding. Firstly, as I will discuss in the next section, evolutionary accounts of salience and conventions (such as Binmore's one) have their own problems. In particular, they fail (or more precisely they give up the intention to) to explain the reasoning process which leads agents to act in some particular way. Instead, they propose a purely ultimate explanation whose relevance for social phenomena might be questioned. Secondly, depending on the decision problem at stake, more or less high orders of mutual knowledge might be necessary for giving a sufficient reason to act. Thirdly, I think Lewis's response to this objection still holds: a theory of successful coordination in repeated strategic decision problems cannot be based on the assumption that individuals entertain false beliefs about each others. Common knowledge and common understanding are then *conceptual* requirements.

https://doi.org/10.1017/S0266267114000261 Published online by Cambridge University Press

following a speech of the Chairman of the Federal Reserve, institutions largely explain why this speech acquires salient properties and figures as a reflexive indicator. Because individuals in the same community share the same institutional heritage, they have confidence in the fact that they interpret the same institutional fact in a similar fashion. They have a common understanding of many situations and they are used to rely on the same institutional signals to make sense of what is happening. Therefore, it is this shared institutional heritage that brings common understanding into a population.

Note that there are two related but separated questions at stake here. First, how is the sharing of reasoning modes (assumption (a)) established at first? Second, how do the members of a given population come to have common knowledge of this sharing in their population? I suggest that these two questions might be answered with a refined version of the concept of *embeddedness*. The sociologist Mark Granovetter (1985) proposed this concept as a remedy to the 'over- and undersocialized conceptions of human action in sociology and economics'. According to Granovetter, the oversocialized conception of human action is pervading in modern sociology. It was developed as a response to the undersocialized conception of the individual that can be found in economics but that actually originates from Thomas Hobbes' political philosophy. Granovetter's embeddedness concept emphasizes the importance of networks and personal relationships in and out of markets for generating trust and discouraging malfeasance (1985: 490). Because economic relations are mostly organized through network structures, agents are actually engaged in repeated interactions where the identity of partners is known. Reputation building is therefore of primary importance and leads to interactions where it is common knowledge that agents have common interests. Granovetter then goes on to argue that social structures, more than morality or institutional arrangements, are the main explanatory device for understanding how trust, coordination and cooperation are generated.

Granovetter's thesis has been heavily discussed and has received some strong criticism, mainly from sociologists. To make embeddedness a useful concept to understand how common understanding and salience emerge in a population, at least two major amendments should be made to Granovetter's analysis. First, Granovetter opposes 'social structures' and 'institutional arrangements'. Given most of the definitions used by economists, it should be clear that institutions are part of social structures. If we identify networks with social structures, we can go further: institutions define networks by imposing rules and norms that regulate social interactions. Second, Granovetter's argument about trust building through personal relationships seems perfectly appealing from a game-theoretic point of view. In fact, repeated games have helped economists to build an important literature on reputation and learning in

https://doi.org/10.1017/S0266267114000261 Published online by Cambridge University Press

strategic interactions. While economic sociology can help to understand the building of network structures generating repeated interactions, embeddedness as such does not bring much insight in the generation of trust. However, in keeping with the idea that embeddedness points to the fact that individual action occurs through network relations and further in communities defined by specific institutions, we have one way to explain how common understanding can be generated.

Of the two questions asked above, the second one, 'How do the members of a given population come to have common knowledge of the sharing of the reasoning modes in their population?' is the most easily answered. It may be reformulated as follows: what is the basis for (Φ - N)-*Common Understanding*? It is the fact that the actions of members of N are embedded into a set of institutions that are constitutive of a *community*. Or, to put it in a shorter form: *the basis for common understanding is community membership*. The idea that community membership is a basis for common knowledge has previously been advanced by Clark and Marshall (1981): 'Even when Ann is not acquainted with Bob, she can assume there are generic and particular things the two of them mutually know. The basic idea is that there are things *everyone* in a community knows and assumes that everyone else in the community knows too' (35, emphasis in original). This echoes Lewis who wrote that 'if a convention, in particular, holds as an item of common knowledge, then to belong to the population in which that convention holds – to be party of it – is to know, in some sense, that it holds' (Lewis 1969: 61). For community membership to be a basis of the common knowledge of a proposition x, two assumptions are required (Clark and Marshall 1981; Sillari 2005): community co-membership and universality of knowledge. The first assumption indicates that being a member of a community entails knowing that everyone else in the population is also a member of it; the second assumption indicates that being a member of the community must entail knowing the proposition x. It is then easy to demonstrate that x must be common knowledge among the members of the community. It should be duly acknowledged that Clark and Marshall (1981) see community membership as a basis for common knowledge, not for common understanding. Community membership is equivalent to a state of affairs in Lewis's account. But as I have repeatedly emphasized here, epistemic assumptions regarding agents' modes of reasoning are needed for a state of affairs to be a basis for common knowledge. Here, I rather suggest that community membership is a reasonable basis for common understanding, i.e. for common knowledge of the sharing of reasoning modes with respect to Φ.

Formally, denote Θ_N as the event that each agent in N is a member of a well-identified community and assume that $\omega \in \Theta_N$. Community membership as a basis for (Φ - N)-*Common Understanding* might be

https://doi.org/10.1017/S0266267114000261 Published online by Cambridge University Press

expressed by the following three axioms:

$$\text{CMA } \Theta_N \Rightarrow K_N(\Theta_N)$$

$$\text{CCM } K_N(\Theta_N) \Rightarrow K_N[K_N(\Theta_N)]$$

$$\text{CBR } K_N(\Theta_N) \Rightarrow K_N(Q) \text{ with } Q = (\Theta - N) - \textit{Shared Reflexivity}$$

The CMA (for *community membership awareness*) axiom simply states that when one is a member of a community, one knows that. The CCM (for *community co-membership*) states that when one knows that he is member of a given community, then one infers that everyone in N knows that all are members of the community. This axiom is grounded on the idea that community membership is more or less a public event because there are many signs that may allow the identification of the community to which one belongs. Finally, the CBR axiom (for *community-based reasoning*) means that the event of knowing that one is a member of the community allows one to infer that shared reflexivity prevails with respect to a set Φ of events (which by definition are salient events). From these three axioms, it follows we can define the *community-based salience* principle:[14]

$$\textit{Community-based salience} : \Theta_N \subseteq K_N^*(Q)$$

The principle of community-based salience indicates that the fact of being a member of a community Θ entails one to assume (Φ - N)-*Common Understanding*. As I have argued, common understanding in a given population defines the set of events that have the properties to generate a common knowledge set of consistent expectations. By definition, these events are *salient*. Therefore, in this framework, community membership is the basis for salience. The first question, 'How is symmetric reasoning established at first?' has still not been answered though. Moreover, the definitions of common understanding and of community-based salience are grounded on the assumption that everyone infers the same regarding symmetric reasoning from community membership. I explore these points in the last two sections where I contrast my account with 'naturalistic' approaches of salience.

6. NATURAL SALIENCE AND 'GILBERT'S CONUNDRUM'

In the past two decades, several game theorists have tried to incorporate salience into a game-theoretic framework. Almost all of these attempts build on the seminal contributions of Schelling and Lewis. Though this is also the case of the community-based salience framework I propose

[14] The proof of this proposition is similar to the proof of Lewis's common knowledge given is section 2. CCM and CBR combined entail $K_N[K_N(Q)]$. Applying CCM to this new proposition adds another level of mutual knowledge, and so on, thus proving common knowledge. Note that agents must take CBR for granted for all community members.

https://doi.org/10.1017/S0266267114000261 Published online by Cambridge University Press

here, most of the alternative accounts are grounded on a fundamentally different conception of salience. I will dub this conception 'natural salience' to contrast it with community-based salience.[15] Because most of the accounts of natural salience start from Lewis's contribution, they tend to suggest that Lewis's own conception of salience was naturalist. Since I am not interested in defending a particular interpretation of Lewis's account, I will not claim that what Lewis actually intended was closer to a community-based conception of salience. I let each reader make up his own opinion on this issue.

To characterize the natural conception of salience is not an easy task because it underlies different types of work using a game-theoretic framework. The constitutive feature of natural salience may be summarized as follows (see also Postema 2008: 43–44):[16] the salience of an entity (an event, a strategy, an outcome) is taken to be a brute, objective and explicit fact. Salience is then a characteristic constitutive of an entity which can be recognized 'straightforwardly' by agents. In some versions of the naturalistic account, this property gives a decisive reason for action, at least when there is no other reason for action. This conception is sometimes explicitly attributed to Schelling and underlies some of the experiments made by Mehta *et al*. (1994). In other versions, salience is still an objective feature but is a tool for coordination only when agents are able to reason on second or higher-order salience, i.e. *i* believes that *x* is salient for *j*. This refined conception of natural salience is for example the one endorsed by cognitive hierarchic theories, e.g. Bardsley *et al*. (2010) or Stahl and Wilson (1995). The common characteristic of these naturalistic versions is that they assume that each agent takes others to be rational to a lesser degree than him. A natural conception of salience is also constitutive of various theoretical studies that introduce the concepts of focal points and salience through strategy labelling or framing (e.g. Sugden 1995; Bacharach 2006). Here the salience of an outcome or a strategy follows from an exogenous set of frames or labelling. Sugden's concept of 'presumptive principal of correct play', which he argues can

[15] For discussions of natural salience accounts, see Hansen and Rojo Arjona (2011) and Postema (2008). Postema's concept of 'public salience' has close connections with my account of community-based salience.

[16] The definition of natural salience and its demarcation with community-based salience is not straightforward. Not all the discussions of salience in a game-theoretic context, including some of the examples I give in the text, endorse an unambiguously naturalistic approach. For some of them, it might be argued that they adopt an understanding of salience akin to my community-based account. I do not reject this possibility though it is a fact that no author has proposed to distinguish several kinds of salience along the lines I suggest. In any case, I do not want to pretend that the distinction is crystal clear and that *all* authors have entertained a naturalistic conception. However, a proper recognition of this distinction is lacking, as well as a formal account of it.

https://doi.org/10.1017/S0266267114000261 Published online by Cambridge University Press

be found in Lewis's work, also displays a naturalistic stance: 'Project regularities in population-wide behaviour, using the analogies that you perceive as most natural, then choose the action that can be expected to produce the best results for you, given that projection' (Sugden 2011: 41).[17] Finally, evolutionary accounts of the emergence of salience, e.g. Binmore and Samuelson (2006) and Alberti *et al*. (2012), also entertain a naturalistic kind of salience. Indeed, in these evolutionary models, an exogenous parameter measuring the degree of salience of various features is introduced.

The various accounts of natural salience all share a distinctive feature: common knowledge (of rationality or of modes of reasoning) is no longer a necessary condition for coordination. This is the clearest in cognitive hierarchic theories since each agent assumes that others reason along a finite (and low) level of orders of thinking. This is also obvious in evolutionary accounts of salience since in this case agents are assumed to behave according to rudimentary learning heuristics. For this reason, it is doubtful that Lewis endorsed the naturalist conception of salience. Somehow, some of the critiques of Lewis's account of salience originate from the failure to recognize this fact. A representative example is Margaret Gilbert's discussion of Lewis's theory of convention (Gilbert 1989, 1990, 1992). Focusing on salience from precedent, Gilbert disputes the claim that it might be a useful coordinating device: '[T]he fact that two agents have in the past done their part in an equilibrium so far says nothing about what either will do in the future' (Gilbert 1990: 12). This 'impotence of precedent' (Gilbert's words) can be generalized to any other kind of salience: 'I conclude that rational agents as such cannot be expected to do their parts in a given salient solution in an otherwise problematic coordination problem' (Gilbert 1992: 336). Gilbert's conclusion is based on the hardly disputable claim that the mere fact that an outcome O is salient to i cannot give i sufficient (practical) reason to play his part in O. This is true even if i suspects that O is also salient to j because j will play his part in O only if he believes that O is salient for i. We then have a thinking process of reduction *ad infinitum* that can give no agent a decisive practical reason to play his part in O. However, it should be noted that Gilbert's argument is grounded on two assumptions: firstly, in Lewis's account, agents are rational in a game theoretic sense;[18] secondly, salience is an objective property of an outcome or of any other

[17] Since Sugden argues that different agents may have different 'private conceptions of salience', it seems that salience is not an objective feature in his analysis. Still, his applied discussion on inductive inference in science clearly displays a naturalistic stance.

[18] By this, Gilbert (1992: 321) means that (a) agents never make a mistake, (b) they always act as reason dictates, (c) they do as well as possible given their preferences.

https://doi.org/10.1017/S0266267114000261 Published online by Cambridge University Press

entity. Thus, Gilbert seems to assume a natural kind of salience.[19] Now, as Gilbert shows, these two assumptions together fail to demonstrate that salience can be a coordination device. Gilbert (1992: 336) then suggests that maybe Lewis was not assuming that agents are fully rational but rather he endowed them with 'brute tendencies to conform to successful precedent'.

Gilbert's critique seems to lead to a conundrum. Since we cannot at the same time assume rational agents (endowed with the ability to form common knowledge expectations) and (natural) salience, it seems that we have to give up one of them. As I have suggested above, naturalistic conceptions of salience give up the rationality assumption. Salience comes from the fact that agents have natural tendencies to discover some regularities or features and to behave according to them. Agents do not really reason about these regularities; they simply follow them. Of course, they reason even less about how others reason about them. The second alternative is endorsed by Gilbert, who simply suggests giving up salience as an explanation of coordination. Recently, some authors have even intended to account for coordination by keeping neither the assumption of rationality nor the concept of salience. A representative example is Brian Skyrms's evolutionary account of conventions and norms (Skyrms 1996).[20] Though he quotes Lewis and borrows his canonical signalling model, Skyrms explains coordination through purely evolutionary models with bounded rational agents. According to him, we simply do not need salience to explain coordination through conventions and norms. Regarding Lewis's account of the origins of language, Skyrms (1996: 102) writes: '[w]here did the requisite common knowledge come from? And where is the salience? … These difficulties disappear if we frame our game theory in terms of evolution rather than in terms of rational decision. Common knowledge is no longer required. Neither is salience.'

7. COMMUNITY-BASED SALIENCE AND NATURAL SALIENCE

So, we may not need salience after all. Like all approaches giving up both salience and common knowledge, evolutionary accounts of coordination allow for insightful rational reconstructions of norms and conventions. They are valuable explanations because they propose

[19] As one referee has noted, whether Gilbert truly entertains a naturalistic account of salience is not easy to establish. This is the same difficulty than the one I indicate in footnote 16 above: there is no explicit discussion of the possibility that several kinds of salience coexist. However, the very fact that Gilbert concludes that rational agents cannot used salience to coordinate is an indication that she is likely to have a naturalistic understanding of salience.

[20] See also Young (1998), Skyrms (2004) and Alexander (2007), for similar arguments.

https://doi.org/10.1017/S0266267114000261 Published online by Cambridge University Press

plausible scenarios through which humans might learn how to coordinate. However, whatever their value is, these approaches do not demonstrate that we *should* completely give up the concept of salience. I see at least three arguments against this idea. Firstly, evolutionary accounts of norms and conventions extensively build on evolutionary game theory whose use in the study of cultural evolution leads to several methodological and theoretical issues. Beyond the fact that the results may be highly sensible to the choice of a particular learning dynamic (which in turn necessitates the grounding of this choice on experimental investigation), the interpretation of payoffs as utilities entails many conceptual difficulties such as making interpersonal comparisons of utility (see Sugden 2002; Grüne-Yanoff 2011). Secondly, evolutionary models only provide 'how possible' types of explanations. In other words, they provide sufficient but not necessary conditions for the evolution of a particular norm. In particular, alternative explanations which are more realistic regarding the reasoning process of agents in decision problems may still be valuable. Finally, Gilbert's conundrum, which may give a justification for a purely evolutionary approach, no longer holds with a community-based conception of salience (Postema 2008).

This last point needs further clarification. Gilbert's conundrum comes from the fact that an objective property of an entity (an event or an outcome) cannot be sufficient for rational agents engaged in a coordination problem to act. Indeed, as we have seen, this leads to an infinite regress thinking. But this is no longer true with community-based salience. The reason for this is threefold: firstly, remember that salience exists in virtue of the fact that there is common understanding with respect to a set of events Φ in a group of agents N, that is (Φ - N)-*Common Understanding*. Secondly, note that the set Φ cannot exist without an already established community where a set of individuals interact within a common institutional framework. In the community-based conception of salience, salience does not give reason for action to members of N (in a causal sense), rather salience is *constituted* by the fact that members of N are members of a community and this is this fact that gives reason for action. The property of salience is to generate common knowledge expectations among the members of the population, and this property comes from common understanding and thus from community membership. Thirdly, in the community-based understanding of salience, the set of salient events Φ is *logically* related to the possibility of coordination. Indeed, the following proposition can be proved (see the appendix for a proof):

> Proposition 2: (Φ - N)-Common Understanding is a necessary and sufficient condition for two or more Bayesian rational players to coordinate on a correlated equilibrium.

https://doi.org/10.1017/S0266267114000261 Published online by Cambridge University Press

This proposition indicates that any correlated equilibrium is generated by the fact that there is (Φ - *N*)-*Common Understanding*. Indeed, under a community-based understanding of salience, a salient event belonging to the set Φ is a sufficient but also a necessary device of coordination between *rational* agents.[21] Gilbert's conundrum is thus solved. This definitely demonstrates the importance of communities and community membership to account for coordination. A great illustration of this framework can be found in Michael Chwe's insightful study of various rituals (Chwe 2001). Chwe provides numerous examples demonstrating how public rituals help to foster common knowledge among a population. He explicitly links the creation of common knowledge and the existence of communities. Though he does not use the concept, Chwe also emphasizes the importance of the common understanding of a situation or of a communication to make common knowledge possible: 'The point here is that common knowledge depends crucially on how each person understands or interprets how other people understand or interpret a communication' (2001: 83, see also 7). This requires that members of a community share a symbol system and a world view, making people symmetric reasoners.

It is worth noting that this account provides an alternative (and maybe complementary) explanation to the 'analogy thesis' that Sugden (2004), following David Hume, endorses to explain the pervasiveness of some conceptions of salience, notably in the case of property conventions. Sugden's basic argument is that in a given situation humans have a propensity to expand through analogy, the frames used in similar but different situations. This analogical reasoning makes salient some of the characteristics of an interaction I' because the participants take it to be analogous to an interaction I and use the same criteria of salience. Sugden's argument relies on a conception of salience as a natural phenomenon. However, according to community-based salience, individuals can use the same conception of salience not because they reason through analogy, but simply because, as members of a community, they assume that they use the same symbol system and the same world view.

Still, some issues remain unresolved. The question asked above, 'How is the sharing of reasoning modes established at first?' has not been answered. The properties that make communities a reliable basis for

[21] I emphasize 'rational' because of course the above proposition is no longer true if agents are endowed with a 'lower' form of rationality. In particular, if agents use some types of heuristics to make decisions without reflecting on others' beliefs or choices, then common understanding is no longer a necessary condition for coordination. This is Gilbert's conundrum: one can eventually accept a naturalistic account of salience but only if he gives up the assumption of full rationality. Of course, the comparative relevance of the naturalistic and the community-based accounts is also an empirical matter.

https://doi.org/10.1017/S0266267114000261 Published online by Cambridge University Press

common understanding have not been uncovered either. I must admit that these are difficult questions for which I do not have straightforward answers. I can however provide some speculations. Regarding the first issue, reasoning modes can be understood as capacities and dispositions which individuals may have acquired either through biological evolution or cultural learning. Kinship and/or group selection at the biological level may possibly explain why members of the same group share the same modes of reasoning and why modes of reasoning are not uniform across groups. But I think that cultural evolution is more likely to explain the fact that different people may not infer the same conclusion from the same set of premises. Reasoning modes, in particular inductive standards, are part of what John Searle (1995) calls the 'Background', i.e. 'the set of non-intentional or pre-intentional capacities that enable intentional states of function' where capacities are 'abilities, dispositions, tendencies, and *causal structures generally*' (1995: 127, emphasis is original). Such capacities are acquired through repeated interactions with specific institutions and through cultural learning. Because members of a community tend to interact essentially with other members, this gives a basis for the convergence of reasoning modes inside a community. A similar speculation can be made for the second issue. Because our evolutionary history (both biological and cultural) has been shaped by the importance of communities, the latter are a reliable basis for common understanding. In some ways, community-based salience works along a form of collective intentionality as characterized by Searle (2010: ch. 3; Hédoin 2013). This is because *we* recognize ourselves as members of the same community that you and I assume that we reason along the same lines.

As this last paragraph suggests, the opposition between natural salience and community-based salience is only partial. Not only some cases of salience may have a naturalistic explanation, but the very possibility of community-based salience has naturalistic foundations. This is not surprising since our social capacities are the product of evolution. What is rejected however is the claim that we can understand salience along purely naturalistic and individualistic lines where salience is a natural and objective property of some entities. Quite the contrary, salience mostly relies on the existence of communities and institutions.

8. CONCLUSION

Building on Lewis's theory of common knowledge, this article defends a community-based account of salience. Though many game theorists consider salience as a useful and relevant concept for solving the problem of indeterminacy, they generally conceive it along naturalistic lines. In the 'natural salience' approach, salience is an objective property of events or outcomes that results from the fact that agents are assumed to share

https://doi.org/10.1017/S0266267114000261 Published online by Cambridge University Press

some basic reasoning modes. Apart from the fact that this may not be the meaning that Lewis was intending when he used the salience concept, both empirical and logical arguments suggest that this not a fully satisfactory account. In particular, what I have called the 'Margaret Gilbert's conundrum' indicates that we cannot retain a naturalistic conception of salience while assuming rational agents. Some authors such as Gilbert or Skyrms simply suggest giving up on the concept of salience. Here, I rather entertain a different conception of it.

Along with the community-based account, I made three claims. Firstly, salience is not always an objective property that gives reason for action in a causal way. Rather, (community-based) salience is constituted by the fact that members of a population have a common understanding of some well-identified events. Common understanding means that individuals have common knowledge of the fact that they share some reasoning modes regarding specific events. Secondly, I suggested that common understanding is very close to the so-called Harsanyi doctrine and common priors assumption in game theory and therefore that salience may be understood through the concept of correlated equilibrium. Finally, I argued that the basis for common understanding is community membership. This last claim can probably be tested experimentally.

Some issues still remain unresolved. The causal factors explaining why members of the same community happened to share the same modes of reasoning are still to be fully uncovered. Why and how community membership serves as a reliable basis for common understanding is not fully understood. Evolutionary and naturalistic theories of culture and institutions may be relevant here. Moreover, it is very likely that in a significant number of cases, salience works along the naturalistic lines traditionally entertained by game theorists. A more empirically based distinction between the two kinds of salience is needed. But the main point made in this article, i.e. that we cannot account for all cases of salience along purely individualistic and naturalistic lines, still holds.

REFERENCES

Alberti, F., R. Sugden and K. Tsutsui. 2012. Salience as an emergent property. *Journal of Economic Behavior and Organization* 82: 379–394.

Alexander, J. M. 2007. *The Structural Evolution of Morality*. Cambridge: Cambridge University Press.

Aumann, R. J. 1976. Agreeing to disagree. *Annals of Statistics* 4: 1236–1239.

Aumann, R. J. 1987. Correlated equilibrium as an expression of Bayesian rationality. *Econometrica* 55: 1–18.

Bacharach, M. 2006. *Beyond Individual Choice: Teams and Frames in Game Theory*. Princeton, NJ: Princeton University Press.

Bardsley, N., J. Mehta, C. Starmer and R. Sugden. 2010. Explaining focal points: cognitive hierarchy theory versus team reasoning. *Economic Journal* 120: 40–79.

Binmore, K. 2008. Do conventions need to be common knowledge? *Topoi* 27: 17–27.

https://doi.org/10.1017/S0266267114000261 Published online by Cambridge University Press

Binmore, K. and L. Samuelson. 2006. The evolution of focal points. *Games and Economic Behavior* 55: 21–42.
Chwe, M. S.-Y. 2001. *Rational Ritual: Culture, Coordination, and Common Knowledge*. Princeton, NY: Princeton University Press.
Clark, H. and C. Marshall. 1981. Definite reference and mutual knowledge. In *Elements of Discourse Understanding*, ed. A. Joshi, B. Webber and I. Sag, 10–63. Cambridge: Cambridge University Press.
Cubitt, R. P. and R. Sugden. 2003. Common knowledge, salience and convention: a reconstruction of David Lewis's game theory. *Economics and Philosophy* 19: 175–210.
Gilbert, M. 1989. Rationality and salience. *Philosophical Studies* 57: 61–77.
Gilbert, M. 1990. Rationality, coordination, and convention. *Synthese* 84: 1–21.
Gilbert, M. 1992. *On Social Facts*. Princeton, NJ: Princeton University Press.
Gintis, H. 2009. *The Bounds Of Reason: Game Theory and the Unification of the Behavioral Sciences*. Princeton, NJ: Princeton University Press.
Granovetter, M. 1985. Economic action and social structure: the problem of embeddedness. *American Journal of Sociology* 91: 481–510.
Grüne-Yanoff, T. 2011. Evolutionary game theory, interpersonal comparisons and natural selection: a dilemma. *Biology and Philosophy* 26: 637–654.
Hansen, P. and D. Rojo Arjona. 2011. Prune or cut down: salience and Sugden's "The Economics of Rights, Cooperation and Welfare". *International Review of Economics* 58: 53–78.
Hargreaves-Heap, S. and Y. Varoufakis. 2004. *Game Theory: A Critical Introduction*. London: Routledge.
Harsanyi, J. C. and R. Selten. 1988. *A General Theory of Equilibrium Selection in Games*. Cambridge, MA: MIT Press.
Hédoin, C. 2013. Collective rationality in economics: making Searle's theory of institutional facts relevant for game theory, *Erasmus Journal for Philosophy and Economics* 6: 1–27.
Hollis, M. and R. Sugden. 1993. Rationality in action. *Mind* 102: 1–35.
Levine, D. 2011. Neuroeconomics. *International Review of Economics* 58: 287–305.
Lewis, D. K. 1969 [2002]. *Convention: A Philosophical Study*. Oxford: Blackwell Publishing.
Mehta, J., C. Starmer and R. Sugden. 1994. The nature of salience: an experimental investigation of pure coordination games. *American Economic Review* 84: 658–673.
Morris, S. 1995. The common prior assumption in economic theory. *Economics and Philosophy* 11: 227–253.
Paternotte, C. 2011. Being realistic about common knowledge: a Lewisian approach. *Synthese* 18: 249–276.
Postema, G. J. 2008. Salience reasoning. *Topoi* 27: 41–55.
Rescorla, M. 2011. Convention. *The Stanford Encyclopedia of Philosophy*, ed. Edward N. Zalta. http://plato.stanford.edu/archives/spr2011/entries/convention/.
Savage, L. J. 1954. *The Foundation of Statistics*. Mineola, NY: Courier Dover Publications.
Schelling, T. C. 1960. *The Strategy of Conflict*. Cambridge, MA: Harvard University Press.
Searle, J. R. 1995. *The Construction of Social Reality*. London: Simon and Schuster.
Searle, J. R. 2010. *Making the Social World: The Structure of Human Civilization*. Oxford: Oxford University Press.
Sillari, G. 2005. A logical framework for convention. *Synthese* 147: 379–400.
Sillari, G. 2008. Common knowledge and convention. *Topoi* 27: 29–39.
Skyrms, B. 1996. *Evolution of the Social Contract*. Cambridge: Cambridge University Press.
Skyrms, B. 2004. *The Stag Hunt and the Evolution of Social Structure*. Cambridge: Cambridge University Press.
Stahl, D. O. and P. W. Wilson. 1995. On players' models of other players: theory and experimental evidence. *Games and Economic Behavior* 10: 218–254.

https://doi.org/10.1017/S0266267114000261 Published online by Cambridge University Press

Sugden, R. 1991. Rational choice: a survey of contributions from economics and philosophy. *Economic Journal* 101: 751–785.
Sugden, R. 1995. A theory of focal points. *Economic Journal*, 105: 533–550.
Sugden, R. 2002. The evolutionary turn in game theory. *Journal of Economic Methodology* 8: 113–130.
Sugden, R. 2004. *The Economics of Rights, Cooperation and Welfare*, 2nd edn. New York, NY: Palgrave Macmillan.
Sugden, R. 2011. Salience, inductive reasoning and the emergence of conventions. *Journal of Economic Behavior and Organization* 79: 35–47.
Vanderschraaf, P. 1998. Knowledge, equilibrium and convention. *Erkenntnis* 49: 337–369.
Vanderschraaf, P. and G. Sillari. 2009. Common knowledge. *Stanford Encyclopedia of Philosophy*, ed. Edward N. Zalta. http://plato.stanford.edu/archives/spr2009/entries/common-knowledge/.
Young, H. P. 1998. *Individual Strategy and Social Structure: An Evolutionary Theory of Institutions*. Princeton, NJ: Princeton University Press.

APPENDIX: PROOFS OF PROPOSITIONS 1 AND 2

This appendix sketches formal proofs for two key propositions made in the main text. Both proofs rely on a theorem established by Robert Aumann (1987) that demonstrates that Bayesian rational players with a common prior p over Ω in an epistemic game G will implement a correlated equilibrium. The converse also holds: for every correlated equilibrium in a game G, one can define an epistemic game G^* with a common prior p over Ω such that in every states ω Bayesian rational players will carry the move corresponding to the correlated equilibrium (Gintis 2009: 136–139). Henceforth, I refer to this result as Aumann's Theorem.

Proposition 1: Equivalence between the common prior assumption and symmetric reasoning

In the fourth section, I argue that the common prior assumption is formally equivalent to assume that players in a game are symmetric reasoners. To prove this claim, it is necessary and sufficient to establish that (1) the common prior assumption implies that players are symmetric reasoners and (2) that when players are symmetric reasoners they in fact have a common prior over Ω.

(1) Assume that Bayesian rational players have a common prior $p(.,\omega)$ over Ω in an epistemic game G. By Aumann's Theorem, the players will implement a correlated equilibrium $(\Omega, f(\omega))$ in a corresponding game G^* where Nature observes the state of the world ω and sends to each player i an instruction $s_i^*(\omega)$. The strategy profile $\mathbf{s^*}(\omega) = (s_1(\omega), \ldots, s_n(\omega))$ is a correlated distribution given by a correlated device f over a probability space (Ω, p) with $f(\omega) = \mathbf{s^*}(\omega)$ for all ω. In G^*, Nature's move corresponds to an event A. Assume that A contains only one element ω $(A = \{\omega\})$ and is public, i.e. each player knows for sure which state of the world ω obtains. Formally, for all i, $A \subseteq K_i(A)$, where $K_i(A) = \{\omega \mid P_i\omega \subseteq A\}$. Assume moreover that each agent i has a single best response $s_i^*(\omega)$. This implies that each player knows the other players' strategies at the equilibrium. Hence, in G^*, all players are able to infer from the publicness of A the same correlated distribution $\mathbf{s^*}(\omega)$ and because A is public each player knows that. Substitute $\mathbf{s^*}(\omega)$ for P to

https://doi.org/10.1017/S0266267114000261 Published online by Cambridge University Press

obtain *Lewis's A symmetric-reasoning condition* as stated in the fifth section of the main text.

(2) Assume that agents are A-symmetric reasoners in a game G. Denote ω as the actual state of the world. Assume a public event A with $\omega \in A$, thus making A common knowledge among the players of G. Denote $\phi_i{}^{\omega}$ the conjectures of player i regarding how other players will play when the actual state of the world is ω. By the very definition of the *Lewis's A-symmetric reasoning condition*, all the players form the same conjectures from A, *i.e.* $\phi_1{}^{\omega} = \ldots = \phi_n{}^{\omega}$ for the n players in G. Since players are Bayesian rational, their conjectures must be consistent, and thus they form a correlated equilibrium. By Aumann's Theorem, this implies that players have a common prior over Ω. □

Proposition 2: (Φ - N)Common Understanding is a necessary and sufficient condition for two or more Bayesian rational players to coordinate on a correlated equilibrium.

We assume that players maximize expected utility and have consistent beliefs. We can prove proposition 2 by contradiction by showing that the following two propositions are inconsistent:

(a) A set N of players coordinate on a correlated equilibria in an epistemic game G.
(b) The set Φ of salient events is empty in G for the members of N.

Assume that (a) is true. By definition, proposition (b) implies that there is not (Φ - N)*Common Understanding* among the players in N. By definition again, this means that members of N are not symmetric reasoners in G. By proposition 1 above, this implies that the members of N do not have a common prior over Ω in G. By Aumann's Theorem, we know that players who implement a correlated equilibrium in G must have a common prior over Ω. Thus, we know that (a) is wrong, which contradicts our initial assumption. This establishes that common understanding is necessary for Bayesian rational players to implement a correlated equilibrium. As for sufficiency, note that (Φ - N)*Common Understanding* implies symmetric reasoning. By proposition 1, we know that symmetric reasoning and the common prior assumption are equivalent. Aumann's Theorem then shows that this must lead players to implement a correlated equilibrium in G. □

https://doi.org/10.1017/S0266267114000261 Published online by Cambridge University Press