The Early History of Common Knowledge

Friday 21st October, 2016
Please don't cite without permission!

Abstract

This essay charts the early history of the notion of common knowledge. The notion was independently developed perhaps as many as nine times in the years 1949-1976. I clear up a number of common confusions about the history of the notion, and suggest that Thomas Schelling's contribution to the development of the idea was more important than many have recognized.

Some people commonly know a proposition just in case they all know it, they all know that they all know it, they all know that they all know that they all know it, and so on. Brought to prominence in the late 1960s and early 1970s, the technical notion of common knowledge has come to play an important role in economic theory, in theoretical computer science, in linguistics, and also in analytic philosophy. The notion has been used in analyses of basic aspects of small-scale interactions such as coordination, joint attention and shared intention, in theories of complex linguistic phenomena such as speaker meaning, conversational context, and the use of innuendo, and even in discussions of interactions on a much larger scale, such as conventions, social groups, and social practices ranging from the use of circular kivas as forums for debate in prehistoric America to the progress of the newly crowned king in medieval Java or England through the lands over which he would rule.[1]

---

[1]Coordination: Lewis (1969), Heal (1978), Moses (1986), Halpern & Moses (1990), which became Fagin *et al.* (1995, Ch. 6, Ch. 11) and Fagin *et al.* (1999), Morris (2002), Morris (2014); Heinemann *et al.* (2004), Thomas *et al.* (2014). Joint attention: Peacocke (2005) (although cf. Campbell (2002, Ch. 8)). Shared intention: e.g. Tuomela & Miller (1988, p. 375), Bratman (1992, p. 335 with n. 15), Bratman (1993, passim, with n. 8), Bratman (2014, p. 5 with n. 9), Fine (2012), cf. Gilbert (1990, p. 3). Linguistic meaning: Schiffer (1972), Grice (1969). Conversational context: Radford (1969) Stalnaker (1978, 1998, 2002, 2014), Clark (1996). Cf. Clark

1

In this essay I chart some aspects of the early history of the technical notion of common knowledge. The notion was independently developed perhaps as many as nine times in the years 1949-1976.

The first reference to the notion that I am aware of occurs in passing. Ruesch & Bateson (1949) proposed methodological foundations for the study of social interactions.[2] They note that there is a particular difficulty when the scientist herself is a participant in the interaction she wishes to study, and is also aware of her role in that interaction. This "participant-observer" "must attempt to be aware of his own knowledge, not only to record it as a part of his observation, but also to see it as a factor which determines his own actions" (111). In passing Ruesch and Bateson consider an objection to this aspect of their theory, namely that it might lead to an "infinite regress": "a method of description within which it could conceivably be relevant to report '*A* knows that *B* knows that *A* knows that *B* knows...that such and such.'" Dubbing this infinite series "fallacious", they reply to the objection by arguing that

> In practice such infinite series or even long series of this kind do not occur, because the participant individuals do not think only in individual terms but also in terms of the interacting system. They are thus able to short-circuit the infinite series by stating implicitly or explicitly '*We* are agreed about such and such' – a statement about the group as a system. The infinite series is an artifact derived from a too rigid adherence to propositions stated only in individual terms. (111)

A decade later, Thomas Schelling initiated the historically most important strand of work on common knowledge. In the penultimate paragraph of Schelling (1959), also published as Schelling (1960a), Schelling writes:

> If the Yalu River is to be viewed as a "limit" in the Korean War that was recognized on both sides, its force and authority is to be analyzed not in terms of the *joint unilateral* recognition of it by both sides in the conflict—not as something that we and the Chinese recognized unilaterally and simultaneously—but as something that we "mutually recognized." It was not just

that we recognized it and they recognized it, but that we recognized that they recognized it, they recognized that we recognized it, we recognized that they recognized it, we recognized that they recognized that we recognized it, and so on. It was a *shared* expectation. To that extent it was a somewhat undeniable expectation; if it commands our attention, and we expect it to be observed, and we expect the Chinese to expect us to observe it, we cannot unilaterally detach our expectations from it. In that sense, limits and precedents and traditions of this kind have an authority that is not granted to them voluntarily by the participants in a conflict; they acquire a magnetism or focal power of their own. (1960a, p. 136-7)

This passage displays two typical features of Schelling's development of ideas related to common knowledge. First, Schelling does not use "know" or "believe" or their cognates explicitly; in the passage above, he uses the word "recognize", so that perhaps he should not be credited with a discovery of common knowledge, but rather with a discovery of "common recognition". Second, Schelling focuses his attention on agents' expectations. His far more influential and widely read *Strategy of Conflict* also exhibits these features. To my knowledge the book does not even contain a discussion of hierarchies of mutual recognition (never mind knowledge or belief), but only discusses hierarchies of expectation. In a coordination problem, he writes,

> The final outcome must be a point from which neither expects the other to retreat; yet the main ingredient of this expectation is what one thinks the other expects the first to expect, and so on...These infinitely reflexive expectations must somehow converge on a single point, at which each expects the other not to expect to be expected to retreat. (Schelling (1960b, p. 70-71, cf. p. 54, 87, *et passim*))

As here Schelling often describes the importance of "mutual expectations" and "the coordination of expectations" to successful coordination of action.

Schelling's emphasis on expectation, given the prevalence of expected utility theory then and now, hardly needs any explanation. But it is worth calling to mind a strikingly parallel passage in von Neumann and Morgenstern's epoch-making book (1944). Although von Neumann and Morgenstern do not discuss Schelling's hierarchy of expectations, they come tan-

3

talizingly close, and their discussion may have been relevant to Schelling's own development of these ideas. In discussing the contrast between Robinson Crusoe's actions when he is alone on an island, and the actions of participants in a social economy, von Neumann and Morgenstern write:

> The difference between Crusoe's perspective and that of a participant in a social economy can also be illustrated in this way: Apart from those variables which his will controls, Crusoe is given a number of data which are "dead"; they are the unalterable physical background of the situation....Not a single datum with which he has to deal reflects another person's will or intention of an economic kind—based on motives of the same nature as his own. A participant in a social exchange economy, on the other hand, faces data of this last type as well: they are the product of other participants' actions and volitions (like prices). His actions will be influenced by his expectation of these, and they in turn reflect the other participants' expectation of his actions.

> Thus the study of the Crusoe economy and the use of the methods applicable to it, is of much more limited value to economic theory than has been assumed heretofore even by the most radical critics. The grounds for this limitation lie not in the field of those social relationships which we have mentioned before—although we do not question their significance—but rather they arise from the conceptual differences between the original (Crusoe's) maximum problem and the more complex problem sketched above.

> We hope that the reader will be convinced by the above that we face here and now a really conceptual—and not merely technical—difficulty. And it is this problem which the theory of " games of strategy" is mainly devised to meet. (11-12)

While the passage in Schelling (1960a) contains what to my mind counts as an explicit discussion of common knowledge, the passages in Schelling (1960b) are much less clear. In what follows, I will speak of those who knew only the latter work as making independent discoveries of common knowledge, although I will continue to remark on the great influence of Schelling in this connection. Even this legislation is somewhat controversial. As noted earlier, Schelling (1960a) only discusses common recognition.

The next explicit discussion of something resembling common knowledge appears to be entirely independent of Schelling's work. Maucorps &

Bassoul (1962) provide an analysis of empathy and its sociological effects and manifestations. They begin with the dialogue of the "ego" and the "alter" (Lat. "other"), which leads to a discussion of the levels of access to knowledge of the other ("paliers d'accès à la connaissance d'autrui"). There are the "expectations of the Self", "expectations of the other", "the other's knowledge of the expectations of the Self" and "the other's knowledge of the self's knowledge of the expectations of the other". The authors recognize the possibly unlimited continuation of this series, but for the purposes of their experiments they restrict attention to the ones described above:

> Si, théoriquement, il n'est pas de limite à un semblable développement, l'expérience montre de façon claire qu'entre le troisième et le quatrième palier la faculté auto-empathique commence a s'enrayer: la difficulté de saisie et l'effort de réflexion entravent l'élan participatif. Pour pallier les effects d'une telle interférence, il convenait de limiter les recherches expermientales aux quatre premiers paliers d'accès à la conscience d'autrui.

> If, theoretically, there is no limit to a similar extension, experience clearly shows that between the third and fourth level the auto-empathetic faculty begins to be checked: the difficulty of grasping and the effort of reflection impede the participant's energy. In order to mitigate the effects of this kind of interference, it was convenient to restrict our experimental studies to the first four levels of access to the knowledge of the other. (48, translation mine)

The remainder of their paper is primarily occupied with an extension of this treatment of empathy to an analysis of the relationship between various social groups: primarily the working-class (le populaire) and outsiders (le exclu), but with some discussion of the very rich (le nanti) and the very poor (le démuni). I've been unable to find subsequent development of these ideas in this tradition of sociology (although see below for discussion of Schütz).

The next three strands in the discussion of common knowledge appear to have been independent of one another, but each of them depends directly on Schelling's work. Two of them (Nozick and Lewis) show knowledge of Schelling (1960b) but not of Schelling (1960a), so according to my convention I will treat them as independent discoveries.

The first of these strands is Robert Nozick's 1963 dissertation, which remained unpublished until 1990. In his prescient discussion, Nozick seeks to

characterize the kinds of real-life situation for which game-theoretic analysis would be appropriate. He does so partly in epistemic terms, defining what he calls a "game-theoretic situation" as one in which there is common knowledge of a variety of background facts, for example, of the payoff structure of the game, of the rationality of the players, of the actions available to each of them, and so on (Nozick, 1990, p. 273-274). In his analysis, Nozick does not use the term "common knowledge" to describe the state, but he does define the notion explicitly. He was aware of Schelling (1960b), but he does not cite Schelling (1960a).

Nozick argues against von Neumann and Morgenstern's minimax solution to zero-sum games (as defended also by Luce & Raiffa (1957)), and in favor of a decision-theoretic approach to the play of games. He notes that common belief in rationality was presupposed, although not made explicit in Harsanyi's early work (Nozick, 1990, p. 284, note)). (I'll return to Harsanyi in a moment.) Nozick does not say whether he takes his criticism of the "maximize security value" decision criterion also to extend to Nash equilibrium, but it seems natural to think that he would have accepted the extension. Nozick's work had, as far as I am aware, almost no impact; it was only much later that his important ideas would come into more mainstream game theory in the founding works of what is now known as "epistemic game theory". But when they did, they did not arise from study of Nozick, but were independent contributions by Aumann, Bernheim, Pearce and those who followed them.

It is worth pausing to reflect on the importance of Harsanyi's work to the history of common knowledge. On a cursory glance, and with the full benefit of hindsight, Harsanyi's work from the late 1950s and early 1960s seems shot through with near-misses of common knowledge and related concepts. For example, in a paper on bargaining between two parties who are ignorant of each other's utility functions, Harsanyi describes a sequence of higher-order estimates (Harsanyi (1957), published as Harsanyi (1962a)). In his model, the parties aim to play the Zeuthen-Nash bargaining solution, but they cannot do so because they do not know the utility function with respect to which the other will calculate this solution. Harsanyi argues that they best they can do is to substitute their *estimate* of the other's utility function for the real utility function in the Zeuthen-Nash solution. But – Harsanyi continues – if the parties think that they will both use estimated utility functions in place of the others' actual utility functions, they should no longer use the estimated utility functions, but instead use their their estimates of the others' estimates of their own utility function. And the same reasoning recurs at higher and higher orders. For if, in turn, they recog-

nize this policy of using the estimates of the estimates, they should use a third-level estimate, and so on *ad infinitum*. They will have well-defined optimal policies only if the relevant series of estimates converge. As Nozick observes, it is not merely that Harsanyi seems close to common knowledge; his formal analyses on occasion even seem to depend informally on a background assumption of something similar to common knowledge of rationality. A similar point could be made about others of Harsanyi's papers on related topics, some of which respond explicitly to Schelling (e.g. Harsanyi (1961, 1962b). But while Harsanyi seems to have been on the verge of defining common knowledge, I have not found in these papers any explicit statement of the notion, or recognition of its importance to the analyses in question.

The second strand of work dependent on Schelling begins with a paper by Thomas Scheff (1967), in which he provides a theory of consensus. He characterizes the notion of complete consensus as follows:

> Complete consensus on an issue exists in a group when there is an infinite series of reciprocating understandings between the members of the group concerning the issue. I know that you know that I know, and so on. This is the definition of *complete* consensus. In actual research, one might find it difficult to locate a single example of such complete consensus, and of demonstrating that it occurred if one did find it. For actual situations, one can derive various degrees of partial consensus, depending upon the level of co-orientation achieved. The zero level would represent agreement, but not consensus.. The first level...would be a first degree consensus, the second level (we recognized that they recognized that we recognized) a second degree consensus, and so on. (37)

Scheff's first concern is to operationalize this "interactionist" alternative to the "individualist" account of consensus in terms of first-order agreement in a population. He presents some modest formalism and discusses the representation of various intuitive cases. He next suggests a method for testing this hypothesis on the basis of the relationship between consensus and coordination; the interactionist model predicts a relationship between coordination and the higher orders of co-orientation (whereas the individualist model does not) (43). Scheff quotes the passage that I quoted above from Schelling (1960a) (he cites Schelling (1959)) nearly in full (36); he also knew of the paper by Maucorps & Bassoul (1962) (36 n. 19).

7

As with Nozick's discussion of Harsanyi, it is worthwhile reflecting on a body of earlier literature known to Scheff which came close to the notion of common knowledge but did not invoke it explicitly. Scheff notes the relationship of his own work to that Schütz (1962) and Garfinkel (1964). Both of these authors were concerned to describe people's shared attitudes to the social world around them. Schütz begins with the suggestion that people's differing perspectives and backgrounds lead to difficulties in constructing common objects in our social world. He suggests that common sense overcomes these difficulties by making certain idealizations. Some of the idealizations he describes might – if made precise in the appropriate way – entail the presence of common knowledge. His discussion comes close to this observation but does not explicitly draw it out:

> In the natural attitude of common-sense thinking in daily life I take it for granted that intelligent fellow-men exist. This implies that the objects of the world are, as a matter of principle, accessible to their knowledge, i.e., either known to them or knowable by them. This I know and take for granted beyond question. But I know also and take for granted that, strictly speaking, the "same" object must mean something different to me and to any of my fellow-men....Common-sense thinking overcomes the differences in individual perspectives...by two basic idealizations:
>
> i) The idealization of the interchangeability of the standpoints: I take it for granted – and assume my fellow-man does the same – that if I change places with him so that his "here" becomes mine, I shall be at the same distance from things and see them with the same typicality as he actually does; moreover, the same things would be in my reach which are actually in his. (The reverse is also true.)
>
> ii) The idealization of the congruency of the system of relevances: Until counterevidence I take it for granted – and assume my fellow-man does the same – that the differences in perspectives originating in our unique biographical situations are irrelevant for the purpose at hand of either of us and that he and I, that "We" assume that both of us have selected and interpreted the actually or potentially common objects and their features in an identical manner or at least an "empirically identical" manner, i.e., one sufficient for all practical purposes.

> It is obvious that both idealizations, that of the interchangeability of the standpoints and that of the congruency of relevances – together constituting the general thesis of *reciprocal perspectives* – are typifying constructs of objects of thought which supersede the thought objects of my and my fellow-man's private experience. By the operation of these constructs of common-sense thinking it is assumed that the sector of the world taken for granted by me is also taken for granted by you, my individual fellow-man, even more, that it is taken for granted by "Us."... Thus, the general thesis of reciprocal perspectives leads to the apprehension of objects and their aspects actually known by me and potentially known by you as everyone's knowledge (12-13)

It is probable that Schütz was influenced in his formulation of this problem and his solution by earlier discussions of related issues in the phenomenological tradition (for example (Merleau-Ponty, 1945, 407-8)).[3]

Harold Garfinkel (1964) subsequently developed the suggestions of Schütz. Garfinkel makes a number of statements which could be interpreted as implying the infinite hierarchy of common knowledge. For example, describing Schütz's notion of a scene "known in common with others" Garfinkel writes:

> According to Schütz, the person assumes, assumes the other person assumes as well, and assumes that as he assumes it of the other person the other person assumes the same for him...(237)

The quotation is followed by an enumeration of the content of these joint assumptions. In reading both this preliminary remark and the enumerated conditions of what it takes for something to be "known in common", one has the sense that we are very close to sufficient conditions for common knowledge. The symmetry condition implicit in the above quotation is especially suggestive. But the consequence is not drawn out explicitly. Scheff rightly contrasts his analysis with that of these earlier authors precisely on the basis of his own interest in yet higher levels of knowledge (37).

Scheff's work inspired a sophisticated, insightful paper by Morris Friedell (1969 with a 1967 working paper). Friedell's paper seems to have been the first to use the term "common knowledge" for the notion. His ambitions were wide ranging: the paper presents new technical results in epistemic

---

[3]Thanks to Taylor Carman for bringing this passage to my attention in a different context.

logic and game theory, as well as new applications for the notion. It is the first to use Hintikka's (1962) approach to epistemic logic in the treatment of common knowledge; Friedell proves a fundamental result which allows for an alternative semantic characterization of common knowledge in this setting.[4] Friedell motivates the result by observing that in the presence of introspection assumptions on individual belief and standard idealizations about agents' logical abilities, if for all p a subject A believes that if A believes that p, then B believes that p, then A will believe they have common belief (a relative of this argument can now be found in Greco (forthcoming)). Later he claims that making eye-contact has the interesting feature that if A sees B looking at A, then B sees A looking at B, and suggests that this explains how eye-contact could give rise to common knowledge. The paper closes with an early application of his machinery of common knowledge to two-person games. His discussion goes far beyond even what is in Nozick in terms of its technical results and the detailed development of the applications. Unfortunately it too appears to have had essentially no impact on the development of the field.

The third important strand of work which is indebted to Schelling (after Nozick and Scheff-Friedell) is the early work of David Lewis. In his 1975 paper "Languages and Language", first drafted in 1968 (Lewis (1975): 6; cf. Lewis (1973)), Lewis explicitly introduces the notion of all knowing, all knowing that all know, and so on. This paper grew out of Lewis's dissertation, in which he defended the notion of meaning by convention against attacks on the notion by his advisor Willard Quine (see e.g. Quine (1936, 1960)). In his earlier work, Lewis had characterized the notion of a convention in part by a the occurrence of a state which he called "common knowledge", but which does not coincide with the technical notion described above. Only in this later paper did he adopt the iterated definition, although he did not here use the name "common knowledge" for it.

It is often mistakenly claimed that Lewis defined common knowledge in the technical sense in *Convention* (1969). There are certainly remarks that suggest he had the general idea in mind, but what Lewis calls "common knowledge" in *Convention* is not equivalent to common knowledge in the

---

[4]Aumann and Lewis are sometimes credited with the theorem that common knowledge is a fixed point; this result was proven rigorously for both common belief and common knowledge as "theorem k" in Friedell (1967, p. 12) (cf. (Friedell, 1969, p. 32, theorem j)), and Friedell himself recognized it as an important result. The development of the understanding of common knowledge as the maximum solution to a particular equation, in the preceding column of the paper is also important; these ideas are often credited to Monderer & Samet (1989).

now standard technical sense.

It is worth remarking that there are also substantial differences in Lewis's use of "common knowledge" in his PhD thesis *Conventions of Language* (1967) and his use of the term in the published version of the thesis, *Convention*. In the thesis Lewis uses "common knowledge" to describe an infinite hierarchy of "potential expectations". After the discussion of the hierarchy of potential expectations, Lewis asks how this hierarchy could be produced, and states something analogous to what would later become his official account of "common knowledge": how some "evidence" *E*, which all "observe", and which is itself evidence that all have observed *E* could generate the hierarchy of potential expectations. In *Convention*, the hierarchy of potential expectations is no longer given its own name, and the name "common knowledge" is now used for something closely related to what was originally the nameless theory of how the infinite hierarchy of potential expectations could be produced. In the book, some agents have "common knowledge" that *p* just in case there is some *E* such that *E* entails *p*, all have reason to believe *E*, and *E* "indicates" to all (as opposed to "being evidence") that all have reason to believe (as opposed to "observe") *E*. Lewis still goes on to claim that this condition (which he now calls "common knowledge") gives rise to something similar to the hierarchy of higher-order expectations.

Both works clearly show the influence of Schelling (1960b) (Lewis does not, as far as I am aware, cite Schelling (1960a)), although Lewis's motivation for introducing the hierarchy of potential expectations is importantly different from Schelling's. Lewis seems to have recognized that higher-order expectations were not in fact required for successful coordination of action (see, e.g. 1969, p. 59). He introduced higher-order expectations simply because he believed the examples of a convention exhibited them, and that these expectations were part of the concept of a convention.

While these differences are of some historical interest, the main point in the present context is that the notions of full-blown common belief and common knowledge do not have center stage in either Lewis's book or his dissertation. In these earlier works, Lewis does claim that if some people have common knowledge that *p* (on his definition of that idea), then under certain special circumstances, they will all believe that *p*, all believe that they all believe that *p*, and so on (this is now known as "common belief"). But Lewis does not explicitly state sufficient conditions under which his definition would give rise to common belief (see Cubitt & Sugden (2003)); at most, the standard technical notions are implicit, below the surface of his discussion. It is interesting that Lewis himself later regretted his choice of

the term "common knowledge" to name the notion defined in *Convention*, since that notion does not even entail that the agents know the relevant claim; something can be "commonly known" in the sense used in *Convention* even if it is false (Lewis, 1978, p. 44 n. 12).

So much for the first round of work which was in one way or another indebted to Schelling. We turn now to a further, entirely independent discovery of the notion, by Colin Radford (1969) – the sixth discovery we have encountered, after Ruesch and Bateson, Schelling, Maucorps and Bassoul, Nozick, and Lewis. In his entertaining, clear paper, Radford argues that ordinary instances of telling, such as when "a man...tells his wife that there is a hole in his cricket socks" require complex iterated knowledge (332). He introduces the idea of an "explicit situation" which accompanies such ordinary cases of telling, and suggests that although these situations are at first sight quite simple, they may require common knowledge. His argument to this conclusion closely resembles an argument later given by Heal (1978) and independently Clark & Marshall (1981), which would become influential in psycholinguistics. He presents a series of putatively non-explicit situations, in which some people know that others know that...something has occurred, with n repetitions of "know", but fail to know...that they know this, with $n + 1$ repetitions of "know". The suggestion is that these cases could be repeated at any finite level; no matter the level at which n iterations failed, the situation would still fail to be explicit. Radford himself does not draw a firm conclusion, though he presents admirably clear discussion of both sides. He did not use the name "common knowledge" for the indefinitely iterated state he defined.

In his 1970 dissertation (published in 1972); Stephen Schiffer used the notion of common knowledge, which he called "mutual knowledge" in the case of two agents. (A 1969 paper by Paul Grice was written in part in response to unpublished work by Schiffer (see Grice (1969, fn. 1)); Grice coined the unsuccessful expression "super-knows" to describe common knowledge.)[5] In the published version, Schiffer cites Schelling (1960b), but like Lewis and Nozick he does not appear to have known Schelling (1960a), so we may count his work as the seventh independent discovery of the notion.

Schiffer developed a particular style of counterexample to Grice's anal-

---

[5]Schiffer (p.c.) recalls developing the notion in September or October of 1967. He reports that Lewis, then teaching at UCLA, often traveled to Berkeley (where Schiffer was teaching), and that at the time, Lewis said he liked Schiffer's name for the state ("mutual knowledge") better than his own. These conversations may have antedated Lewis's use of the now standard definition of common knowledge in his 1968 draft of Lewis (1975).

ysis of speaker meaning, and introduced the notion of mutual knowledge to avoid similar counterexamples to his own account. A simple version of Grice's analysis held that a speaker means that *p* by an utterance if and only if the speaker believes that the utterance will cause the hearer to believe that *p* because of the hearer's belief that the speaker intended the hearer to come to believe that *p* on the basis of the utterance. Schiffer's counterexamples targeted this analysis; he proposed cases in which speaker and hearer each believed that the other believed that they had the relevant intention, but believed that the other did not believe that they themselves believed that they had the intention. Schiffer suggested that this kind of example could be produced no matter how many finite iterations of "all believe that" or "all know that" were added to the analysis; he introduced the idea of common knowledge as a way of escaping all counterexamples of this form.

At around the same time, Robert Stalnaker (1970, 1973, 1974) and Lauri Karttunen (1973, 1974, Karttunen & Peters (1975)) began to develop the notion of speaker presupposition and common ground. The phrases "common knowledge", "presumed common knowledge" and even "common belief" are used in Stalnaker (1973) and Stalnaker (1974), but the formal definition of common knowledge was not given in those papers, and some remarks suggest that the phrases are used in a more ordinary English sense; Stalnaker (1978) first clearly identified the common ground with the presumptive common knowledge (in the technical sense) of the participants in a conversation.

Robert Aumann (1976) subsequently discovered the notion an eighth time. Aumann's paper was also the third independent use of the phrase "common knowledge" to describe the notion. Aumann of course knew of Schelling's work by this time, but he does not seem to have known Schelling (1960a). It is sometimes claimed that Aumann attributed the notion to Lewis, but he does not cite Lewis's book in his paper. In a recent interview, however, Aumann himself disclaims the independence of his own discovery.[6] Aumann recalls the story with characteristic verve, and it is worth quoting his remarks in full:

> The David Lewis matter is really very, very interesting. I'll tell you my side of the story.
>
> I wrote this paper and called this concept 'common knowledge'. This was published in 1976. Now around 1979 I ran across a paper in a philosophical journal which quoted my paper, I think.

---

[6]Thanks to Joe Halpern and Robert Aumann for bringing this interview to my attention.

I'm not sure, though. It certainly quoted Lewis's book *Convention*, which was published in '69. Perhaps it discussed a citation from my paper, too: I don't remember now, but it certainly quoted Lewis's book. I opened my eyes and said: 'Hey! What's going on here? This chap had the concept of common knowledge already in '69?' And, yes, it turned out that he did! I went back and bought the book or took it out of the library, and there it was. In 1969! And the amazing thing was that we used the same word for it, the same word that I thought I had invented. So from then on, when I wrote about this I started quoting Lewis.

Now, a year or two later Lewis sent a letter to the provost of Princeton University. In his PhD thesis, Sergio Ribeira DaCosta Werlang had quoted me on common knowledge. Somehow the thesis was passed on to David Lewis as a reader. Lewis was furious because the thesis didn't quote him at all. It made believe that common knowledge was my concept. So I think Lewis wrote a letter to the provost in which he said...well, he complained about this in fairly strong language. I don't remember the exact language. He sent a copy to Hugo Sonnenschein and a copy to me. So I got this copy and immediately wrote back saying: 'you know, Prof. Lewis, this is your concept. No question about it. I was not aware of your contribution to this when writing the '76 paper. I simply did not know. You might find this difficult to believe because we use the same word, we use the word 'common knowledge'. But it's true. And as soon as I became aware, which is a year or two ago now, I started citing you. Please accept my abject apologies. I'm not even saying that this is independent work. You can talk about independent work when you're talking about something that is done at approximately the same time or even a year or two later, but not when there's a hiatus of seven years. Because, you know, people talk at lunch or they talk in seminars and things filter through, very often without attribution. Both the idea itself and the name could have filtered through somehow without my being aware of it. So I'm not claiming independence. It's your contribution. The idea of common knowledge is your contribution and there's no question about it. This is all yours. I've been saying it for a while now, and I'll say it again. And let's meet.'

I was in Stony Brook at that time. I went down to Princeton,

and I met Lewis and Kripke. We had a nice conversation. By that time there was a kosher dining club in Princeton. I'll have to tell you some stories about Princeton when I was originally there in the fifties. It was a hotbed of anti-Semitism. But you know, the world progresses and by 1981-82 there was a kosher dining club and in fact Hugo Sonnenschein became provost of Princeton. He's Jewish, of course. And then afterwards there was even a Shapiro who became President of Princeton University. He is also Jewish. So we're making progress. I had a nice day in Princeton and made it up with Lewis and Kripke, and everything was fine. And at every opportunity, including this one, I've been acknowledging that the idea of common knowledge is Lewis's idea. Period. No independence, nothing.

However, let me just add this: Lewis had the idea of common knowledge, and this is very clear. What he did not have is the agreement theorem. He did not have the agreement theorem, he had no glimpse of the agreement theorem, nothing like that. So I still lay claim to the agreement theorem. (Aumann & Roy (2010, 24-5); cf. Hart (2005, 698))

Aumann's agreement theorem is a striking and elegant result about common knowledge in a Bayesian setting. It says roughly that if two ideal agents have common knowledge of what probability they assign to a given proposition, then they they assign that proposition the same probability. This result was the first intimation that, together with other assumptions about epistemic rationality, one could provide simple and elegant analyses relating the epistemic states of many agents. The result became the basis of a number of further important results (e.g. the "no trade" theorem of Milgrom & Stokey (1982)), developed in Aumann's signature "common prior" setting (for discussion see Gul (1998), Aumann (1998)).

In the fifteen years following Aumann's paper, in part following up on the insight of the agreement theorem, common knowledge came to the forefront of a branch of game theory. To mention just some of the important papers in this period: Bernheim (1984), Pearce (1984), Aumann (1987), Brandenburger & Dekel (1987), Monderer & Samet (1989), Rubinstein (1989) (an early survey was Geanakoplos (1994); Dekel & Siniscalchi (2015) provide a more recent one). In computer science early discussions came from Akkoyunlu *et al.* (1975), Gray (1978), Cohen & Yemini (1979) and Yemini & Cohen (1979). These early papers in computer science did not formalize the notion of knowledge explicitly, and so did not represent com-

mon knowledge explicitly. Major advances in this direction were made by Yoram Moses under the supervision of Joseph Halpern (Moses (1986)). Parts of this work were published as Halpern & Moses (1990), which can now be found in the important textbook Fagin *et al.* (1995).

Barwise (1989, p. 203) reports that John McCarthy had developed the notion of common knowledge around 1960, but McCarthy's collected papers (McCarthy & Lifschitz (1990)) do not show any evidence of this. Correspondence with some who knew McCarthy and his work fairly intimately has not unearthed any further direct evidence of McCarthy's early work on common knowledge (as opposed to "common sense"). The paper Barwise cites, McCarthy *et al.* (1978), uses an operator ("every fool knows...") which implies every finite level of "all agents know that..."; perhaps it should be listed as a ninth independent discovery.

References

Akkoyunlu, Eralp A., Ekanadham, Kattamuri, & Huber, R. V. 1975. Some Constraints and Tradeoffs in the Design of Network Communications. *SIGOPS Oper. Syst. Rev.*, **9**(5), 67–74.

Aumann, Robert J. 1976. Agreeing to Disagree. *The Annals of Statistics*, **4**(6), 1236–1239.

Aumann, Robert J. 1987. Correlated Equilibrium as an Expression of Bayesian Rationality. *Econometrica*, **55**(1), 1–18.

Aumann, Robert J. 1998. Common Priors: A Reply to Gul. *Econometrica*, **66**(4), 929–938.

Aumann, Robert J., & Roy, Olivier. 2010. Robert Aumann. *Pages 21–33 of: Epistemic Logic: 5 Questions. Vincent F. Hendricks and Olivier Roy, eds*. Automatic Press.

Barwise, Jon. 1989. *The situation in logic*. Center for the Study of Language (CSLI).

Bernheim, B. Douglas. 1984. Rationalizable Strategic Behavior. *Econometrica*, **52**(4), 1007–1028.

Brandenburger, Adam, & Dekel, Eddie. 1987. Rationalizability and correlated equilibria. *Econometrica*, **55**(6), 1391–1402.

Bratman, Michael E. 1992. Shared cooperative activity. *The philosophical review*, 327–341.

Bratman, Michael E. 1993. Shared intention. *Ethics*, 97–113.

Bratman, Michael E. 2014. *Shared agency: A planning theory of acting together*. Oxford University Press.

Campbell, John. 2002. *Reference and consciousness*. Oxford University Press.

Chwe, Michael Suk-Young. 2001. *Rational Ritual: Culture, Coordination, and Common Knowledge*. Princeton University Press.

Clark, Herbert H. 1979. Responding to indirect speech acts. *Cognitive psychology*, **11**(4), 430–477.

Clark, Herbert H. 1996. *Using language*. Cambridge University Press.

Clark, HH, & Marshall, CR. 1981. Definite reference and mutual knowledge. 10–64.

Cohen, Danny, & Yemini, Yechiam. 1979. Protocols for Dating Coordination. *Pages 179–188 of: Proceedings of the Fourth Berkeley Conference on Distributed Data Management and Computer Networks*.

Cubitt, Robin P, & Sugden, Robert. 2003. Common knowledge, salience and convention: A reconstruction of David Lewis' Game Theory. *Economics and Philosophy*, **19**(2), 175–210.

Dekel, Eddie, & Siniscalchi, Marciano. 2015. Epistemic Game Theory. *Pages 619–702 of:* Young, H. Peyton, & Zamir, Shmuel (eds), *Handbook of Game Theory with Economic Applications, Volume 4.*

Fagin, Ronald, Halpern, Joseph Y., Moses, Yoram, & Vardi, Moshe Y. 1995. *Reasoning about Knowledge.* MIT Press.

Fagin, Ronald, Halpern, Joseph Y., Moses, Yoram, & Vardi, Moshe Y. 1999. Common knowledge revisited. *Annals of Pure and Applied Logic*, **96**(1 - 3), 89 – 105.

Fine, Kit. 2012. *The structure of Joint Intention.*

Friedell, Morris. 1967 (April). *On the structure of shared awareness.* Working Paper, available: http://deepblue.lib.umich.edu/bitstream/handle/2027.42/50812/27.pdf?sequence=1.

Friedell, Morris F. 1969. On the structure of shared awareness. *Behavioral Science*, **14**(1), 28–39.

Garfinkel, Harold. 1964. Studies of the routine grounds of everyday activities. *Social problems*, **11**(3), 225–250.

Geanakoplos, John. 1994. Common Knowledge. *Chap. 40, pages 1437–1496 of:* Aumann, Robert J., & Hart, Sergiu (eds), *Handbook of Game Theory with Economic Applications*, vol. 2. North Holland.

Gilbert, Margaret. 1989. *On social facts.* Princeton University Press.

Gilbert, Margaret. 1990. Walking together: A paradigmatic social phenomenon. *MidWest studies in philosophy*, **15**(1), 1–14.

Gray, Jim. 1978. Notes on Data Base Operating Systems. *Pages 393–481 of: Operating Systems, An Advanced Course.* London, UK: Springer-Verlag.

Greco, Daniel. forthcoming. Iteration Principles in Epistemology I: Arguments For. *Philosophy Compass.*

Grice, H. P. 1969. Utterer's Meaning and Intention. *The Philosophical Review*, **78**(2), pp. 147–177.

Gul, Faruk. 1998. A Comment on Aumann's Bayesian View. *Econometrica*, **66**(4), 923–927.

Halpern, Joseph Y, & Moses, Yoram. 1990. Knowledge and common knowledge in a distributed environment. *Journal of the ACM (JACM)*, **37**(3), 549–587.

Harsanyi, John C. 1957 (December 11). *Bargaining in ignorance of the opponent's utility function.* Cowles Working Paper No. 46.

Harsanyi, John C. 1961. On the rationality postulates underlying the theory of cooperative games. *Journal of Conflict Resolution*, 179–196.

Harsanyi, John C. 1962a. Bargaining in ignorance of the opponent's utility function. *Journal of Conflict Resolution*, 29–38.

Harsanyi, John C. 1962b. Rationality postulates for bargaining solutions in cooperative and in non-cooperative games. *Management Science*, **9**(1), 141–153.

Hart, Sergiu. 2005. An Interview with Robert Aumann. *Macroeconomic Dynamics*, **9**(05), 683–740.

Heal, Jane. 1978. Common knowledge. *The Philosophical Quarterly*, **28**(111), 116–131.

Heinemann, Frank, Nagel, Rosemarie, & Ockenfels, Peter. 2004. The Theory of Global Games on Test: Experimental Analysis of Coordination Games with Public and Private Information. *Econometrica*, **72**(5), 1583–1599.

Hintikka, Jaako. 1962. *Knowledge and Belief.* Cornell University Press.

Karttunen, Lauri. 1973. Presuppositions of compound sentences. *Linguistic inquiry*, 169–193.

Karttunen, Lauri. 1974. Presupposition and linguistic context. *Theoretical linguistics*, **1**(1), 181–194.

Karttunen, Lauri, & Peters, Stanley. 1975. Conversational Implicature in Montague Grammar. *In: Annual Meeting of the Berkeley Linguistics Society*, vol. 1.

Lee, James J, & Pinker, Steven. 2010. Rationales for indirect speech: the theory of the strategic speaker. *Psychological review*, **117**(3), 785.

Lewis, David. 1973. Lingue e lingua. *Versus*, **4**, 2–21.

Lewis, David. 1975. Languages and language. *Minnesota Studies in the Philosophy of Science*.

Lewis, David. 1978. Truth in fiction. *American Philosophical Quarterly*, 37–46.

Lewis, David K. 1969. *Convention: A philosophical study*. Harvard University Press.

Lewis, David Kellogg. 1967. *Conventions of Language*. Ph.D. thesis, Harvard University.

Luce, Robert Duncan, & Raiffa, Howard. 1957. *Games and Decisions: Introduction and Critical Survey*. Courier Corporation.

Maucorps, P-H, & Bassoul, René. 1962. Jeux de miroirs et sociologie de la connaissance d'autrui. *Cahiers internationaux de Sociologie*, **32**, 43–60.

McCarthy, John, & Lifschitz, Vladimir. 1990. *Formalizing common sense: papers*. Vol. 5. Intellect Books.

McCarthy, John, Sato, Masahiko, Hayashi, Takeshi, & Igarashi, Shigeru. 1978. *On the model theory of knowledge*. DTIC Document, No. STAN-CS-78-657. Stanford University Department of Computer Science.

Merleau-Ponty, Maurice. 1945. Phénoménologie de la perception. *Paris: Gallimard*, 381.

Milgrom, Paul, & Stokey, Nancy. 1982. Information, trade and common knowledge. *Journal of Economic Theory*, **26**(1), 17–27.

Monderer, Dov, & Samet, Dov. 1989. Approximating Common Knowledge with Common Beliefs. *Games and Economic Behavior*, **1**(2), 170–190.

Morris, Stephen. 2002. Coordination, Communication, and Common Knowledge: A Retrospective on the Electronic-mail Game. *Oxford Review of Economic Policy*, **18**(4), 433–445.

Morris, Stephen. 2014. Coordination, timing and common knowledge. *Research in Economics*, **68**(4), 306–314.

Moses, Yoram O. 1986. *Knowledge in a distributed environment*. Ph.D. thesis, Stanford University.

Nozick, Robert. 1990. *The normative theory of individual choice*. New York: Garland Books.

Peacocke, Christopher. 2005. Joint attention: Its nature, reflexivity, and relation to common knowledge. *In:* Eilan, Naomi, Hoerl, Christoph, McCormack, Teresa, & Roessler, Johannes (eds), *Joint Attention: Communication and Other Minds*. Oxford: Oxford University Press.

Pearce, David G. 1984. Rationalizable Strategic Behavior and the Problem of Perfection. *Econometrica*, **52**(4), 1029–1050.

Pinker, Steven. 2007. The evolutionary social psychology of off-record indirect speech acts. *Intercultural pragmatics*, **4**(4), 437–461.

Pinker, Steven, Nowak, Martin A., & Lee, James J. 2008. The logic of indirect speech. *Proceedings of the National Academy of Sciences*, **105**(3), 833–838.

Quine, Willard V. 1936. Truth by convention. *In: Philosophical Essays for A. N. Whitehead*. Longmans, Green and Co.

Quine, Willard Van. 1960. Carnap and logical truth. *Synthese*, **12**(4), 350–374.

Radford, Colin. 1969. Knowing and Telling. *The Philosophical Review*, **78**(3), pp. 326–336.

Rubinstein, Ariel. 1989. The Electronic Mail Game: Strategic Behavior Under "Almost Common Knowledge". *The American Economic Review*, **79**(3), 385–391.

Ruesch, Jurgen, & Bateson, Gregory. 1949. Structure and process in social relations. *Psychiatry*, **12**(2), 105–124.

Scheff, Thomas J. 1967. Toward a sociological model of consensus. *American Sociological Review*, 32–46.

Schelling, Thomas C. 1959. *Toward a Theory of Strategy for International Conflict*. Rand.

Schelling, Thomas C. 1960a. The Retarded Science of International Strategy. *Midwest Journal of Political Science*, **4**(2), 107–137.

Schelling, Thomas C. 1960b. *The strategy of conflict*. Harvard University Press.

Schiffer, Stephen R. 1972. *Meaning*. Oxford: Oxford University Press.

Schütz, Alfred. 1962. *Collected papers, 3 vols*. The Hague: Martinus Nijhoff.

Stalnaker, Robert. 1998. On the representation of context. *Journal of Logic, Language and Information*, **7**(1), 3–19.

Stalnaker, Robert. 2002. Common ground. *Linguistics and Philosophy*, **25**(5), 701–721.

Stalnaker, Robert C. 1970. Pragmatics. *Synthese*, **22**, 272–289.

Stalnaker, Robert C. 1973. Presuppositions. *Journal of philosophical logic*, **2**(4), 447–457.

Stalnaker, Robert C. 1974. Pragmatic presuppositions. New York: New York University Press.

Stalnaker, Robert C. 1978. Assertion. *Pages 315–32 of:* Cole, P. (ed), *Syntax and Semantics*, vol. 9. New York Academic Press.

Stalnaker, Robert C. 2014. *Context*. Oxford: Oxford University Press.

Thomas, Kyle A, DeScioli, Peter, Haque, Omar Sultan, & Pinker, Steven. 2014. The psychology of coordination and common knowledge. *Journal of personality and social psychology*, **107**(4), 657.

Tuomela, Raimo, & Miller, Kaarlo. 1988. We-intentions. *Philosophical Studies*, **53**(3), 367–389.

von Neumann, John, & Morgenstern, Oskar. 1944. *Theory of Games and Economic Behavior*. Princeton University Press.

Yemini, Yechiam, & Cohen, D. 1979. Some issues in distributed processes communication. *Pages 199–203 of: Proc. of the 1st International Conf. on Distributed Computing Systems*.