Common Knowledge*

Harvey Lederman

Monday 30th January, 2017

1 Two Notions of Common Knowledge

Common knowledge is widely used in economic theory, theoretical computer science, linguistics, and philosophy. It has been invoked in explanations of rational coordination, in characterizations of joint attention and shared intention, in theories of communication, speaker meaning, innuendo, and the context of conversations, as well as in analyses of conventions, social norms, and social groups.[1]

The expression "common knowledge" in fact has two, distinct, technical uses.[2] Jane Heal introduces the "informal" notion of common knowledge by way of the following case:

*Thanks to Marija Jankovic, Kirk Ludwig, Cédric Patternotte and especially Dan Greco for comments on this article. Thanks also to Peter Fritz for discussion of an earlier version of some of this material.

[1]For coordination, see Lewis (1969), Heal (1978), Moses (1986), Halpern & Moses (1990), Fagin *et al.* (1995, Ch. 6, Ch. 11), Fagin *et al.* (1999), Morris (2002), Morris (2014), Heinemann *et al.* (2004), Thomas *et al.* (2014). For joint attention, see Peacocke (2005) (although cf. Campbell (2002, Ch. 8)). For shared intention, see 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). For communication and speaker meaning, see Strawson (1964), Grice (1969), Schiffer (1972), Sperber *et al.* (1986), and Jankovic (2014). For innuendo, see Clark (1979), Pinker (2007), Pinker *et al.* (2008), Lee & Pinker (2010). For conversational context, see Radford (1969), Stalnaker (1978, 1998, 2002, 2014), Clark (1996), cf. Kölbel (2011). For conventions, see Lewis (1969, 1975). For social norms, see Bicchieri & Muldoon (2014) and references therein. For social groups, see Gilbert (1989). A number of concrete examples of social practices are discussed in Chwe (2001) (see also Friedell (1969)).

[2]Neither of these technical senses is supposed to explicate the ordinary language uses of the expression "common knowledge", for example, the sense in which facts in an encyclopedia are common knowledge.

> Suppose that you and I are dining together. At the next table, clearly audible to both of us, another diner begins a loud quarrel with his companion. You catch my eye and make a grimace of distaste. Contrast this with another situation. You and I are again dining together. I messily drop a piece of potato on the table. I attempt to field it inconspicuously and hope you have not noticed. In fact you have noticed, but out of regard for my feelings you pretend that you have not.
>
> There is a clear difference between the two situations. In the first the beginning of the quarrel is completely open, public, or, as I shall say, it is *common knowledge* between us that a quarrel has broken out. In the second case my bungling is not similarly open. (1978, p. 116)

A second, "formal" notion of common knowledge is defined as follows. Some people *commonly know* that p if and only if they all know that p, they all know that they all know that p, they all know that they all know that they all know that p, and so on. Related notions can be defined using "believe" and "is certain that", as well as other propositional attitudes. For example, some people commonly believe that p if and only if they all believe that p, they all believe that they all believe that p, and so on. Often, especially in economic theory (and almost without exception in economic theory from the 1980s and 1990s), the expression "common knowledge" is used generically to mean "common knowledge, common belief, or common certainty" (where someone is certain that p just in case they assign p probability 1). Thus authors speak of "common knowledge assumptions" where the assumptions may not require knowledge at all. I will sometimes use "common knowledge" in this more general way to simplify the presentation, but where it is important I will mark distinctions between common knowledge, common belief, and common certainty.

To avoid confusion, I will use the expression "public information" for the first, "informal" notion of common knowledge introduced by the example from Heal.[3] From here on, I will use the term "common knowledge" only for the second, formal notion. The reader should be aware, however, that what I call "public information" is often called "common knowledge" by other authors.

[3] This is also intended as a technical term. For example, I take it that "public information" as it used to describe corporate disclosures is not synonymous with "public information" in this sense.

These two different notions give rise to different questions. It is reasonable to ask (as Heal does) whether one can give a clear psychological characterization of public information, and if so, what that characterization is. These questions do not arise for (the formal notion of) common knowledge, since this already comes with a clear psychological characterization. But common knowledge comes with its own set of questions. Do people ever have it? Even if they do, does it play an important role in human social interactions?

While the two notions of common knowledge are conceptually distinct, the default position in the literature has been that they coincide. According to this *Default Position*, some people have public information that p just in case they have common knowledge that p.

This article focuses on conceptual issues related to this Default Position. In Section 2, I consider an influential argument in favor of the position. In Section 3, I consider an influential argument against it. In Section 4, I discuss how one might use a famous formal example, the electronic mail game, to argue for or against the position.

I will not discuss the mathematics of common knowledge, or its uses in game theory and computer science in any detail. Excellent surveys of these issues can be found in other places. Geanakoplos (1994) is accessible and highly recommended. Vanderschraaf & Sillari (2014) touches on many issues which I cannot discuss here. For those wishing to delve deeper into the logic of common knowledge, the excellent textbook Fagin *et al.* (1995) remains the best starting point; for those specifically interested in game theory, Dekel & Siniscalchi (2015) is a useful recent survey. I have not made any attempt to be systematic in my references in this article.

2 An Argument for the Default Position

An influential argument for the Default Position proceeds by way of a series of examples; I'll use those from the recent paper of Greco (2015):[4]

> PUBLIC ANNOUNCEMENT: A professor tells her class that they will play the following game. Without communicating to one another in any way, each student in the class will write down the name of a US state on a piece of a paper. If all students write the

[4] I will present the argument using the notion of public information, although Greco himself does not. Early versions of this argument can be found in Heal (1978) and Clark & Marshall (1981). Throughout this section I focus on common knowledge, but all of the considerations I describe could also be applied, *mutatis mutandis* to common belief.

> same state name, with the exception of the name of the state the class is taking place in, the students will each receive $10. If any two students write down different state names, or if they all write down the name of the state the class is taking place in, no prize money will be awarded. Before handing out the pieces of paper, the professor tells the class that she grew up in Maine (which is not the state the class is taking place in), and that it is lovely in the fall. (Greco (2015, p. 755))

Contrast this case with the following:

> PRIVATE INFORMATION: Just like the previous case, except instead of publicly announcing that she grew up in Maine, the professor whispers the following to each student privately as she hands out the pieces of paper: "while I'm not telling anybody else this, I'd like you to know that I grew up in Maine, and it is lovely in the fall." (Greco (2015, p. 756))

In the first case, the fact that the professor grew up in Maine has the "openness" or "publicity" that Heal describes in her example; the students have public information that the professor grew up in Maine. In the second case, the fact that the professor grew up in Maine is not "open" in the same way; the students do not have public information of this fact.

A theory of public information should explain this contrast. Prior to reading the above examples on might have been attracted to the simple view that some people have public information that p just in case they all know that p. But PRIVATE INFORMATION is a counterexample to this simple theory. In PRIVATE INFORMATION, the students all know that the professor is from Maine. But intuitively this fact is not "open" to them.

In response to this counterexample, one might propose adding a second "layer" of knowledge: some people have public information that p just in case they all know that they all know (two occurrences of "know") that p. But the following elaboration of PRIVATE INFORMATION appears to be a counterexample to this new proposal, as well:

> MORE PRIVATE INFORMATION: Just like the previous case, except this is what the professor whispers: "I'm privately telling everybody in the class that I grew up in Maine and that it's lovely in the fall. However, you're the only one who I'm telling that I'm telling everyone. Each other student thinks that she's

the only one who knows that I grew up in Maine." (Greco (2015, p. 756))

Here, the students presumably all know that they all know that the professor is from Maine. But nevertheless there is a felt difference between MORE PRIVATE INFORMATION and PUBLIC ANNOUNCEMENT: in the former, the students do not seem to have public information that the professor is from Maine.

To consider generalizations of this style of argument, it will be useful to have some terminology. Some people *mutually know* (or: mutually know^1) that p just in case they all know that p. Since everyone in my department knows that the department office is on the tenth floor, the members of my department mutually know^1 that the department office is on the tenth floor. Progressing further, some people mutually know^2 that p just in case they all know that they all know that p. Since everyone in my department knows that everyone in my department knows that the department office is on the tenth floor, the members of my department mutually know^2 that the department office is on the tenth floor. And we can continue: some people mutually know^3 that p just in case they all know that they all know that they all know that p. Since everyone in my department knows that everyone in my department knows that everyone in my department knows that the department office is on the tenth floor, the members of my department mutually know^3 that the department office is on the tenth floor.

Extending this pattern, in general some people mutually know^n that p just in case they mutually know that they mutually know^{n-1} it.[5] We can use these definitions to give a compact characterization of common knowledge: some people commonly know that p just in case for all n, they mutually know^n that p.

The examples given so far seem to show that people may have mutual knowledge^2 that p, although they do not have public information that p. But many authors have agreed that the examples can be extended to show in general that for any n, mutual knowledge^n does not suffice for public information. To give a sense of how the generalization proceeds, we can give a related example for mutual knowledge^3 by altering the teacher's announcement:

EVEN MORE PRIVATE INFORMATION: Just like the previous case,

[5] In older work, e.g. Schiffer (1972), "mutual knowledge" is sometimes used synonymously with "common knowledge". But the terminology in the main text has now become standard; see, e.g. Fagin *et al.* (1995).

except this is what the professor whispers: "I'm privately telling everybody in the class that I grew up in Maine and that it's lovely in the fall and that I have told everyone this. However, you're the only one who I'm telling that I'm telling everyone that I've told everyone. Each other student thinks that she's the only one who knows that everyone knows that I grew up in Maine."

The basic recipe should now be clear. For each n, one can alter the professor's secret so that it seems the students would be able to achieve mutual knowledge^n that the professor is from Maine, although they would not intuitively have public information that the professor is from Maine. The argument based on these cases is thus in the first instance a negative one. A natural analysis of public information states that (for some fixed k) some people have public information that p just in case they have mutual knowledge^k that p. The argument casts doubt on analyses of this form: it appears to show that for any n, no matter how great, mutual knowledge^n is insufficient for public information.[6]

This negative argument can be resisted. The technical notion of public information was introduced on the basis of examples such as Heal's; the notion does not come with hard and fast rules for how it is to be extended to cases such as EVEN MORE PRIVATE INFORMATION. It does not make sense to appeal to "intuitions" here. Rather, in thinking about how to extend the notion we must consider which way of extending it fits best with the theoretical work we want the notion to do. It is at least possible that the most theoretically fruitful extension of this concept gives the result that the people in EVEN MORE PUBLIC INFORMATION (or some later case in the sequence) do have public information, at least if all goes well.

But there might seem to be strong reasons for not extending the concept in this way. For if we did so extend it, we might seem to be left without a way of explaining the undeniable felt contrast between PUBLIC ANNOUNCEMENT and the subsequent cases. This concern, however, is not as powerful as it might seem. To see this, consider a proponent of MK^4, the view that people have public information that p just in case they have mutual knowledge^4 that p. In response to an elaboration of EVEN MORE PRIVATE INFORMATION

[6]In the cases I have taken from Greco, if the students have mutual knowledge^n that the professor is from Maine, they also falsely believe that they do not have mutual knowledge^n that the professor is from Maine. Having the announcement generate this additional false belief simplifies the presentation (and strengthens the intuition), but it is inessential. We could alter this aspect of the teacher's announcement, while preserving the intuitive difference between PUBLIC ANNOUNCEMENT and the other cases.

intended to be a counterexample to MK^4, the proponent of this view might argue as follows: "The students will have mutual knowledge^4 if, and (let us suppose) only if, they (a) understand the secret told to them, (b) believe what the teacher says, (c) know that everyone else will understand the secret, and (d) know that everyone else will believe them. Let us set aside (b) and (d): for present purposes we may grant that if the students understood the speeches they would believe them. Still, (a) and (c) may present difficulties. People in the students' shoes will not always understand the complex statements they are told about others' epistemic states (e.g. 'I told the others I was telling everyone that I was telling everyone that I'm from Maine.'). This would result in a failure of (a). More often, a person in the students' position may be uncertain whether on hearing 'I told everyone that I told everyone that I am from Maine', the *others* will understand that all the students know that all the students know that the teacher is from Maine (a failure of (c)). The difficulty of understanding statements of this form has been documented in the empirical literature.[7] Readers of these cases typically believe that the students do not satisfy (a) or (c), and thus they typically judge that the students do not have mutual knowledge^4 that the professor is from Maine. It is for this reason that readers of the cases judge the students not to have public information of this fact. The example is thus not a counterexample at all; the judgments readers report are exactly in line with my theory." This proponent of MK^4 might further claim that if one focuses on cases where the students do satisfy (a)-(d), the felt contrast disappears. Or she might propose a kind of error theory, stating that the difficulty of satisfying (a)-(d) makes it natural to judge incorrectly that the students fail to have public information in the case so described.

This style of response to the argument in fact has considerable appeal. But it has not been much discussed in the literature, so from now on we will set it aside.[8] In the remainder of this essay, then, we will suppose that the negative argument succeeds. How can this negative argument be extended to a positive argument for the claim that public information requires common knowledge? The most obvious way of doing so uses a further premise: that public information in this case is to be analyzed solely in terms of what the

[7] In Kinderman *et al.* (1998), Stiller & Dunbar (2007), people were found to be essentially at chance in remembering reports involving six nested attitudes. So even if the reader is unhappy with this response to this case, it becomes even more plausible for (e.g.) mutual knowledge^7.

[8] For a different kind of argument against the view that people have public information that p just in case they have mutual knowledge^k that p, see Heal (1978, p. 125).

students know about where the professor comes from, and what they know about what others know about where the professor comes from, and so on. Given this additional premise, if we can rule out all finite levels of mutual knowledge, we will have shown that common knowledge is required for public information.

One might, however, reject the new premise of this extended, positive argument. The sense of openness or publicity in the examples might derive from features of the situation other than how many levels of mutual knowledge are present. For example, situations might fail to produce a sense of "openness" simply when they are confusing, unusual or unfamiliar to the participants. The above cases involve very unusual statements on the part of the professor, and that is why (according to the suggested criterion) that they do not produce the relevant kind of openness. A full development of this approach would require a more systematic explanation of what these other factors are; "unusual" is hardly an informative characterization. But the general strategy for rejecting this new premise again seems at least *prima facie* appealing, and again has not received much discussion.[9]

3 Problems with the Default Position?

Although these arguments have been taken by many authors to motivate the idea that common knowledge is at least importantly related to public information, some of the same authors have also been suspicious that public information requires common knowledge itself. The most common concern has been that because common knowledge requires knowledge of an infinite collection of claims it is impossible for finite creatures to have common knowl-

[9] This alternative strategy might be supported by the suggestion that the examples above prove too much: one can construct an elaboration of them that might suggest that even common knowledge would be insufficient for public information:

> SUPERTASK SECRETS: As in EVEN MORE PRIVATE INFORMATION, but the teacher programs a computer to broadcast the message to the students, and the computer and students are all capable of supertasks. At first the students receive a message which simply says that the teacher is from Maine. After this, at intervals of $\frac{1}{2^n}$ seconds each student's screen displays a new line, saying: $\ulcorner$the others knows that together you have mutual knowledge^n that the teacher is from Maine, but does not know that you know this$\urcorner$.

This case is hard to think about it. But some may be inclined to claim that, although the super-students have common knowledge that the professor is from Maine, they do not have public information of this fact.

edge. Heal (1978) and Clark & Marshall (1981) are good examples of early skepticism on these grounds; Clark for example has written: "CG-iterated [that is, common belief] obviously cannot represent people's mental states because it requires an infinitely large mental capacity..." (1996, 95-6).

In this passage, Clark is concerned with how we should "represent common ground". In context, it is clear that by "represent" he means "give a theory of", and that "common ground" is used here loosely for public information. The objection is not particularly powerful. People often believe every member of an infinite collection of claims. For every natural number $n > 0$, I believe that no one has ever seen Santa Claus exactly n times. This belief about how many times Santa Claus has been seen is clearly possible, in spite of the fact that it involves believing an infinite set of claims. So observing that common knowledge or belief is somehow infinite does not seem to be an interesting objection to the possibility of common knowledge or belief.

Nevertheless, in response to this and related objections, there has been a great deal of discussion of alternative analyses of public information which do not require that people know an infinite collection of claims. Although these analyses do not require that people in fact have common knowledge, they do imply that ideal agents would have common knowledge were they to have public information. In other words, they imply:

IDEAL COMMON KNOWLEDGE: Necessarily, if some agents have common knowledge that they are ideal reasoners, then if they have public information that p, they commonly know that p.[10]

The Default Position is the benchmark position in the literature; it is almost invariably the starting point of discussions about common knowledge and public information. But few people accept the Default Position as it stands. By contrast, almost every author who has discussed public information at any length accepts IDEAL COMMON KNOWLEDGE (or a related thesis about common belief).

An important example of such a theory can be motivated using the notion of "the scene". In the opening example from Heal, there is some sense in which the diners both know the scene which is unfolding around them. The obvious features of what they can see and hear – the color of the table cloth, the location of the bar in the restaurant, whether there is music playing or not – are all part of the scene. Not every detail, of course, must be part

[10] Margaret Gilbert comes closest to advocating this thesis in so many words; Gilbert (1989, p. 186-197) cf. Gilbert (2008)).

of the scene: details of the ornamentation of the ceiling or the color of the waiters' shoes may not be part of the scene in this sense. But it is supposed to be part of the scene that both diners are normal perceivers attentive to the world around them. In other words, it is part of the scene that they each know the scene.

What does it mean to be "part of the scene"? One explication of this abstract idea states that for something to be part of the scene is for the scene to entail it. Thus we have:

> SHARED ENVIRONMENT: Some people have public information that p just in case there is an E ("environment") such that it is the case that E, E entails that p, and E entails that everyone knows that E.[11]

Under plausible assumptions about entailment, and assuming that ideal knowledge is closed under entailment, this definition will satisfy IDEAL COMMON KNOWLEDGE. Informally, the basic idea is as follows. The scene entails that there is a quarrel happening beside the diners. The scene also entails that both diners know the scene. Thus the scene entails that both diners know something which entails that both diners know that there is a quarrel beside them. But since the scene also entails that both know the scene, it further entails that both know something which entails that both know something which entails that both know that there is a quarrel. And we can continue to add iterations of "know something which entails" *ad infinitum.* Given that ideal agents' knowledge is closed under entailment, in the case of ideal agents we can replace each occurrence of "know something which entails" with "know". The formal details of an argument to this effect are sketched in a footnote.[12]

[11] This approach is most closely associated with the work of Barwise (1988); cf. Mertens & Zamir (1985). Further definitions of a related kind may be found in: Lewis (1969) (cf. Cubitt & Sugden (2003) and Paternotte (2011)); Harman (1977, p. 422) (cf. Harman (1974, p. 225), Peacocke (2005, p. 307-8).); Heal (1978); Milgrom (1981), and Clark & Marshall (1981), Clark & Carlson (1982). Lismont & Mongin (2003) gives a clear, simple presentation of some definitions in this family.

[12] For simplicity I use a language that allows quantification into sentence position, and take "entails that" (somewhat unnaturally) to be a binary sentential operator. The argument is not particularly delicate and could be conducted in many other settings, for example, using only first-order quantification, a notion of propositional truth, and entailment understood as a relation between propositions.

First some definitions: q is *mutually entailed by what some agents know* if there is a p such that they mutually know that p, and p entails that q; q is mutually^n entailed by what some agents know if there is a p such that they mutually know p, and p entails that q is mutually^{n-1} entailed by what they know. Finally q is commonly entailed by what some

Earlier I suggested that at least one motivation for these alternatives to common knowledge itself is not as strong as it has been taken to be. But there other ways of using these alternative analyses of public information. My infinite collection of beliefs about Santa Claus is plausibly in some sense explained by my having a more basic belief, namely, that no one has ever seen Santa Claus. What is the corresponding "basic belief" in the case of common belief? If this talk of "basic beliefs" is cashed out in terms of how beliefs are

---

agents know just in case for all n, q is mutually^n entailed by what they know.

We can show that if some people satisfy SHARED ENVIRONMENT with regard to p, then p is commonly entailed by what they know, using four additional assumptions about entailment. The first assumption relies on the idea that propositions are closed under a countable conjunction operation: ENTAILMENT CONJUNCTION-IN: if p entails that q_a for every $q_a \in \{q_a\}_{a \in A}$ where A is countable, then p entails that $\bigwedge_{a \in A} q_a$. Second, DETACHMENT: if q and q entails that p, then p. Third, ENTAILMENT OF ENTAILMENT: if p entails that q, then r entails that [p entails that q]. Fourth, EXISTENTIAL INTRODUCTION: if p entails that $\varphi(q)$, where q is a sentence variable, then p entails that $\exists r \varphi(r)$, where $\exists$ is a quantifier for the type of sentences, and r is a variable of the corresponding type. This fourth principle allows quantifying in, in sentence position: if p entails that S knows q, then p entails that there is something S knows.

Given the above laws, we first prove a lemma: if p entails that everyone knows that q, and if q entails that s, then p entails that there is an r such that [everyone knows that r, and r entails that s]. The proof first uses ENTAILMENT OF ENTAILMENT, then ENTAILMENT CONJUNCTION-IN, and finally EXISTENTIAL INTRODUCTION.

Now, given SHARED ENVIRONMENT, and supposing that some agents have public information that p, there is an E such that E entails that everyone knows E and such that E entails that p. So by the lemma, E entails that there is an r such that everyone knows that r and such that r entails that p. That's the base case of an induction: E entails that p is mutually^1 entailed by what everyone knows. Now for the induction hypothesis, we suppose that E entails that p is mutually^n entailed by what all know, and show that E entails that p is mutually^{n+1} entailed by what all know. Since E entails that all know that E, and by hypothesis E entails that p is mutually^n entailed by what they know, then again by the lemma, E entails that there is some r such that they all know r and r entails that p is mutually^n entailed by what all know. So, by definition E entails that p is mutually^{n+1} entailed by what everyone knows. This induction shows that for all n, E entails that p is mutually^n entailed by what all know, so by ENTAILMENT CONJUNCTION-IN, E entails that p is commonly entailed by what they know. Given that the agents in fact have public information that p, it is the case that E, so by DETACHMENT p is commonly entailed by what they know.

If the agents in question are ideal, and ideal knowledge is closed under entailment (in the sense that if p entails that q and the agent knows that p then the agent knows that q), the same argument can be used to show that they have common knowledge that p, and not merely that p is commonly entailed by what they know. Perhaps a more plausible assumption, that ideal knowledge is closed only under known entailments, would require different background assumptions; the point here has just been to give a worked example of how this kind of view implies IDEAL COMMON KNOWLEDGE.

cognitively represented, answering this question would mean offering a theory of how common knowledge and common belief are cognitively represented. Perhaps SHARED ENVIRONMENT could be suitably modified to provide such an account.

While some have worried that common belief would be too hard to achieve, others have argued that in fact common belief is easier to achieve than mutual belief^k for only finite k. Robert Stalnaker, for example, writes:

> Instead of thinking of beliefs and presuppositions in terms of sentences that express them, in the language of thought, stored in the belief or presupposition box of the mind, think of them negatively: it is the live options – the space of possibilities one allows for – that need to be represented. One's beliefs or presuppositions are the propositions that are true in all of those possibilities. What would put an unrealistic computational load on the mind would be a situation in which one believed, up to six iterations, that one believed that, but failed to believe the seventh iteration. That would require representing a possible situation (and a representation of the complex relational structure of possibilities compatible with possibilities that are compatible with possibilities compatible with...etc.) in which this mind-bogglingly subtle distinction is made. Infinitely iterated beliefs and presuppositions are much simpler for the subject to represent, and to understand. (Stalnaker (2009, 402-3))

Stalnaker's own idea of representation is somewhat obscure. It is not clear in what sense if any including "more" possibilities requires some "computational load". For every n greater than or equal to zero, it is consistent with what I believe that there are exactly n creatures living in the universe now on planets other than earth: an infinite set of live possibilities which differ with regard to the numbers of extraterrestrial creatures are really consistent with my beliefs. Thus making these possibilities live cannot have imposed an "unrealistic" computational load on my mind.

4 Common Knowledge and Coordination

In section 2, we saw an argument that mutual knowledge^n is not sufficient for public information. There, we were focused on the sense of openness that was supposed to be the hallmark of public information. But one might also see the examples as the basis for a different argument, about rational action. In PUBLIC ANNOUNCEMENT, it is rational – perhaps rationally required – for the students to write down "Maine", and it is likely the students would coordinate to win the prize money. In PRIVATE INFORMATION (and also in at least some of the subsequent cases), this behavior does not seem rationally required, and, moreover, it is unlikely that the students would be able to coordinate on writing down "Maine". One theoretical role for public information might be to explain rational social behavior. For example, one might think that in a range of situations it is rational for people to coordinate if and only if they have public information of relevant facts about their situation. If the examples show that finite levels of mutual knowledge are insufficient to justify coordination, we might have a second, independent argument for the Default Position.

But one might wonder: is it really true that the students would still fail to coordinate in the four-hundredth case in this series? It might seem that, whatever one's views about the "openness" of the professor's birthplace in these examples, the students would eventually be able to coordinate. It is quite difficult to think about rational behavior in such alien examples, so it is natural to turn to formal, mathematical theories of rationality for help. The famous "electronic mail game" (Rubinstein (1989)) is a formal example in which it is irrational for players to coordinate if they have any finite level of mutual knowledge, although it would be rational for them to coordinate in the presence of common knowledge. At least in this example, rational students would not coordinate even in the analogue of the four-hundredth case in the series. I'll present the example in very crude outline; readers who want to understand the details should consult the original, highly accessible paper.[13]

In the electronic mail game, Row and Column are uncertain which of two smaller games, G_A or G_B, they are playing (see Figure 4.1). In the first, the action (A) ensures a payoff of 0 to each player, regardless of what the other player does. In the second, there is no such simple choice: the players each receive a payoff of 1 if they both play B, and a payoff of 0 if they both play A. If their actions don't match, however, then the player who plays B alone pays a penalty of 2.[14]

[13] Chant & Ernst (2008) and Lederman (forthcominga) give further discussion of the philosophical import of arguments about the relationship between common knowledge and coordination.

[14] I present a variant of the original electronic mail game, which appeared in Morris (2002). Rubinstein used a slightly different game and obtained a slightly weaker result.

Column (game G_A)

		A	B
Row	A	$0, 0$	$0, -2$
	B	$-2, 0$	$-2, -2$

G_A

Column (game G_B)

		A	B
Row	A	$0, 0$	$0, -2$
	B	$-2, 0$	$1, 1$

G_B

Figure 4.1 The Electronic Mail Game

The games are to be selected by the toss of a fair coin: with probability $\frac{1}{2}$, the game is G_A, and with probability $\frac{1}{2}$ the game is G_B. But the players' information about the outcome of the toss is asymmetric. Row will observe the outcome of the toss. Column will not, but will learn about the outcome by receiving messages from Row. If the game is selected to be G_B, Row will automatically send an email message to Column. If either Column or Row receives an email message from the other, his or her computer automatically sends a new message in reply. At each stage of this process, however, there is a positive, equal and independent rate of the message failing; with probability 1 the process will be cut off at some point. And when it has ended, each player will be told only the tally of the messages he or she sent.

A *strategy* for a player is a function from the natural numbers (the number of messages the player has sent) to the set of probability distributions over actions (the set $\{A, B\}$). A strategy is *rationalizable* if and only if it is consistent with the players' having common certainty of one another's rationality (in the sense of maximizing expected utility). We then have the following result:

Theorem 4.1 (Rubinstein (1989))**.** *For each player, the unique rationalizable strategy is the constant function which takes every number of messages sent to A.*

In the example, coordination is impossible if people have only finite levels of mutual certainty. Many have found this result highly counterintuitive. After one message has been sent, each of the players is certain that the game is G_B – that is, they are certain that if they both played B, that would be the best outcome for each of them. But, because they are not certain that they are certain that the game is G_B, they are certain that the other player will *not* play B. The point carries over to higher message counts, too. If they each send two messages, they are certain that the game is G_B, and in fact certain that the other is certain of this. But even so they remain certain that neither will play B.

In philosophy, perhaps the most common lesson that has been drawn from this result is that common knowledge (or at least: common certainty) is important for successful coordination. If the players had common knowledge that the game is G_B, it would clearly be rationally permitted for them to coordinate on B. But if common knowledge fails in the way described in the electronic mail game, they cannot rationally coordinate on B. So, some have concluded, common knowledge must be important for coordination in general.

There are, however, a number of ways of resisting this argument. Although people cannot coordinate on the better outcome *in the above situation* it does not follow that common knowledge is important for rational coordination in other situations. This point can be sharpened once we see that the argument is sensitive to a number of highly unrealistic features of the setup. For example, Rubinstein already recognized that if it is guaranteed that communication will stop at some pre-ordained number of messages, it becomes possible for players to coordinate rationally on B (Rubinstein (1989, Remark 1, p. 388)). Since in everyday life, communication is always truncated in this way, the example is crucially different from any situation people would ordinarily encounter. A detailed study of consequences of this important observation can be found in the elegant paper of Binmore & Samuelson (2001).

In fact, on reflection, one might take the example to tell against the importance of common knowledge, and not in favor of it. The background assumptions for the above result (including common knowledge of rationality) imply that in the standard technical sense of "rational" the only rational action is to coordinate on A in this situation. But there is a powerful intuition that this is not the only rational action. Instead of taking this case to show that rational agents cannot coordinate, we might instead conclude that the assumptions used in proving the result make the wrong predictions about the case, so that these background assumptions should be rejected. On this interpretation, the case yields an argument against the assumptions of common certainty used to prove the result, not an argument for the role of common certainty in everyday coordination (Lederman (forthcominga)).

One might take this to suggest that any doxastic state resembling common knowledge would impose overly strong demands on when agents could coordinate. But this would be too quick. A person p-believes that q just in case the person assigns q probability greater than or equal to p. It can be shown that for particular choices of p, agents may have common p-belief about the setup of the electronic mail game and common p-belief that they are rational but nevertheless coordinate after finite numbers of messages have been sent (Monderer & Samet (1989)). Thus one might take the argument to show merely that the important notion for theories of behavior is common p-belief, as opposed to common certainty (or common knowledge, or common belief).[15]

We have seen arguments both for and against the prevalence of common knowledge based on the electronic mail game. Other arguments related to this issue are based on the use of common knowledge assumptions in the social sciences more generally. Common knowledge assumptions can be seen as part of simple, strong theories of human behavior. On the grounds that one should generally put confidence in simple, strong hypotheses, one might hold that there is at least a presumption in favor of the claim that people often have a great deal of common knowledge. In reply to this argument, one might argue that any such presumption is defeated by apparent counterexamples to the truth of common knowledge assumptions, such as the false predictions these hypotheses seem to generate in examples such as the electronic mail game itself.

A different argument begins from the claim that the prevalence of models which impose common knowledge assumptions in mathematical game theory shows these assumptions to be part of the best social-scientific theory of human behavior. If one holds that what a person believes and knows is what he or she believes and knows according to the best theory of his or her behavior, then it would follow that people have a great deal of common knowledge and belief.[16] One could resist this argument by rejecting the theory of belief it is based on, or (as above) by rejecting the premise that theories which include common knowledge assumptions are in fact good theories of human behavior.

[15] Note that the standard technical notion of "common p-belief" is in fact defined by analogy to SHARED ENVIRONMENT, and not by analogy to the "iterated" definition given at the start of this article. These definitions coincide if $p = 1$, but they come apart if $p < 1$. For discussion of these differences see Morris (1999). Paternotte (2016) is a recent discussion of differences between common belief and common p-belief, given one popular way of relating belief and p-belief.

[16] For an argument of this kind, see Greco (2014a,b).

5 Conclusion

This brief survey has left untouched a vast range of issues related to common knowledge. Even among the issues it has touched on, many questions remain unsettled. There has not been enough philosophical work on the underpinnings of social behavior to determine whether common knowledge or even public information plays the roles it has been claimed to play. The orthodoxy that common knowledge is importantly connected to public information has not been subjected to sustained scrutiny. As a result there have been few arguments for or against this orthodoxy. A start has been made in some recent work arguing against IDEAL COMMON KNOWLEDGE (Lederman (forthcomingb)), but this remains only a first step.

A good example of the uncertain state of play concerns one of the most important applications of common knowledge and public information in philosophy: the common ground of a conversation. In linguistics and the philosophy of language, it is typically assumed that there is a body of shared information among the participants in a conversation which plays an important role in determining what sentences will mean in context, and how extra-semantic inferences will be drawn by conversational participants. It is standardly assumed that this body of shared information is determined roughly by what the participants commonly know (Stalnaker (2002, 2014), cf. e.g. Pinker (2007), Pinker *et al.* (2008)). There have been few arguments offered in favor of this position, however. It is unclear whether the linguistic data require this assumption, or whether we can make do with something weaker. If we can make do with something weaker, it is unclear what shape this weaker characterization will take. There is a great deal of work to be done.

References

Barwise, Jon. 1988. Three views of common knowledge. *Pages 365–379 of: Proceedings of the 2nd Conference on Theoretical Aspects of Reasoning about Knowledge*. Morgan Kaufmann Publishers Inc.

Bicchieri, Cristina, & Muldoon, Ryan. 2014. Social Norms. *In:* Zalta, Edward N. (ed), *The Stanford Encyclopedia of Philosophy*, spring 2014 edn. Metaphysics Research Lab, Stanford University.

Binmore, Ken, & Samuelson, Larry. 2001. Coordinated action in the electronic mail game. *Games and Economic Behavior*, **35**(1), 6–30.

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.

Chant, Sara Rachel, & Ernst, Zachary. 2008. Epistemic conditions for collective action. *Mind*, **117**(467), 549–573.

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

Clark, H. H., & Marshall, C. R. 1981. Definite reference and mutual knowledge. *Pages 10–63 of:* Joshi, A. K., Webber, B., & Sag, I. (eds), *Elements of discourse understanding.* Cambridge: Cambridge 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, Herbert H, & Carlson, Thomas B. 1982. Speech acts and hearer's beliefs. *In:* Smith, Neilson Voyne (ed), *Mutual knowledge.* Academic Press.

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

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.

Gilbert, Margaret. 2008. Social convention revisited. *Topoi*, **27**(1-2), 5–16.

Greco, Daniel. 2014a. Could KK be OK? *Journal of Philosophy*, **111**(4), 169–97.

Greco, Daniel. 2014b. Iteration and Fragmentation. *Philosophy and Phenomenological Research*, **88**(1), 656–673.

Greco, Daniel. 2015. Iteration Principles in Epistemology I: Arguments For. *Philosophy Compass*, **10**(11), 754–764.

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

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

Harman, Gilbert. 1974. Review of *Meaning* by S. Schiffer. *Journal of Philosophy*, **71**, 224–9.

Harman, Gilbert. 1977. Review of Jonathan Bennett's Linguistic Behaviour. *Language*, **53**(2), 417–424.

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.

Jankovic, Marija. 2014. Communication and shared information. *Philosophical Studies*, **169**(3), 489–508.

Kinderman, Peter, Dunbar, Robin, & Bentall, Richard P. 1998. Theory-of-mind deficits and causal attributions. *British Journal of Psychology*, **89**(2), 191–204.

Kölbel, Max. 2011. What is Assertion? *In:* Brown, Jessica, & Cappelen, Herman (eds), *Assertion: New philosophical essays*. Oxford University Press.

Lederman, Harvey. forthcominga. Two Paradoxes of Common Knowledge: Coordinated Attack and the Electronic Mail Game. *Noûs*.

Lederman, Harvey. forthcomingb. Uncommon Knowledge. *Mind*.

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

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

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

Lismont, Luc, & Mongin, Philippe. 2003. Strong completeness theorems for weak logics of common belief. *Journal of Philosophical Logic*, **32**(2), 115–137.

Mertens, Jean-Francois, & Zamir, Shmuel. 1985. Formulation of Bayesian analysis for games with incomplete information. *International Journal of Game Theory*, **14**(1), 1–29.

Milgrom, Paul. 1981. An axiomatic characterization of common knowledge. *Econometrica: Journal of the Econometric Society*, 219–222.

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

Morris, Stephen. 1999. Approximate common knowledge revisited. *International Journal of Game Theory*, **28**(3), 385–408.

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.

Paternotte, Cedric. 2011. Being realistic about common knowledge: a Lewisian approach. *Synthese*, **183**(2), 249–276.

Paternotte, Cédric. 2016. The fragility of common knowledge. *Erkenntnis*, 1–22.

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.

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.

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.

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

Sperber, Dan, Wilson, Deirdre, , & . 1986. *Relevance: Communication and cognition*. Vol. 142. Citeseer.

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

Stalnaker, Robert C. 2009. On Hawthorne and Magidor on Assertion, Context, and Epistemic Accessibility. *Mind*, **118**(470), 399–409.

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

Stiller, James, & Dunbar, Robin IM. 2007. Perspective-taking and memory capacity predict social network size. *Social Networks*, **29**(1), 93–104.

Strawson, P. F. 1964. Intention and Convention in Speech Acts. *The Philosophical Review*, **73**(4), pp. 439–460.

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.

Vanderschraaf, Peter, & Sillari, Giacomo. 2014. Common Knowledge. *In: The Stanford Encyclopedia of Philosophy*, vol. Spring 2014. Edward N. Zalta.