Who's afraid of common knowledge?*

Giorgio Sbardolini

ILLC and Philosophy Department, University of Amsterdam

February 7, 2024

Abstract

Some arguments against the assumption that ordinary people may share common knowledge are sound. The apparent cost of such arguments is the rejection of scientific theories that appeal to common knowledge. My proposal is to accept the arguments without rejecting the theories. On my proposal, common knowledge is shared by ideally rational people, who are not just mathematically simple versions of ordinary people. They are qualitatively different from us, and theorizing about them does not lead to predictions about our behavior. Nevertheless, models of action that assume common knowledge have a role to play in our understanding of collective rationality.

1 Public information and common knowledge

Something is *common knowledge* in a group if and only if everyone in the group knows it, everyone in the group knows that everyone in the group knows it, everyone in the group knows that everyone in the group knows that everyone in the group knows it, and so on. A similarly defined attitude is common belief. Common attitudes such as common belief and common knowledge are widely used in the study of collective rational action in economics (Aumann, 1976; Osborne and Rubinstein, 1994), computer science (Meyer and van der Hoek, 1995), linguistics (Stalnaker, 2002, 2014), rational choice theory (Fagin et al., 1995), and philosophy (Lewis, 1969; Schiffer, 1972).

A typical application of the common attitudes is in explaining successful coordination. Suppose that two partners, a and b, have decided to meet at an agreed upon location. Neither wants to be there alone: a will go only if a knows that b will go, and b will go only if b knows that a will go. Hence, a will go only if a knows that b knows that a will go. Likewise for b. Hence, a will go only if a knows that b knows that a knows that b knows that a will go. And so on. An infinite regress appears to follow. However, a and b's coordination problem is solved if we assume that a and b have common knowledge that at least one of them will go.

*Acknowledgements: I would like to thank Fausto Carcassi, Andre Curtis Trudel, John Dougherty, Angelica Hill, Alexander Reutlinger, and Sébastien Rivat, as well as audiences at the ILLC/PHILMAT Seminar, and at the 2023 meeting of the BSPS in Bristol. Special thanks to Stephan Hartmann, Luca Incurvati, and Harvey Lederman, for feedback on different versions of this paper. This work has received funding under the EU Horizon 2020 Program within the project *EXPRESS: From the Expression of Disagreement to New Foundations for Expressivist Semantics* (ERC Grant Agreement No. 758540), and under the Horizon Europe Program within the project *Evolution of Logic* (MSCA Grant Agreement No. 101064835).

The lesson appears to be that common knowledge is "needed" to explain coordination (De Freitas et al., 2019, p. 13751), and that it is "a prerequisite for day-to-day activities of coordination and agreement" (Fagin et al., 1995, p. 454). However, the claim that ordinary people have common knowledge, or something analogous to it, is controversial (Heal, 1978; Gilbert, 1989; Sperber and Wilson, 1995; Parikh, 2005; Lederman, 2018b).[1]

To avoid repetitions, I will only mention common knowledge from here on, although much of the discussion below applies to common belief as well, and similarly defined attitudes. I will begin with a review of some reasons to be skeptical of the claim that ordinary people may have common knowledge, and focus in particular on a recent argument by Harvey Lederman (2018b), and on a somewhat similar argument that can be extracted from Robert Aumann (1976). I will refer to these arguments as *skeptical arguments*, since they purport to deny that we ever have common knowledge. (Of course, any skeptic of individual knowledge will deny common knowledge as well, but individual knowledge skepticism is not my concern here.)

Reacting to skeptics of common knowledge, Dan Greco (2014a, p. 170) says: "If these arguments are sound, we must reject explanations of linguistic, economic, and other social phenomena that appeal to common knowledge". This is a high cost, and Greco (among others) rejects the skeptical arguments. I propose instead the following. Some of the arguments against the idea that ordinary people may have common knowledge are sound. However, we need not reject theories in linguistics, economics, or other fields, that appeal to it. Indeed, we need not even revise such theories. The reason is that these theories are not about ordinary people: they are about ideally rational people, who may share common knowledge and for whom the skeptical arguments fail. Since I think that some skeptical arguments against common knowledge are sound, I disagree with Greco, and since I think that there are people for whom the skeptical arguments fail, I disagree with Lederman. My position rests on a proper understanding of the peculiar role played by common knowledge, and by the ideally rational people who might have it, in theories of collective rationality.

2 Arguments against Common Knowledge

According to Fagin et al. (1995), common knowledge is "a prerequisite for day-to-day activities of coordination and agreement" (p. 454). For Vanderschraaf and Sillari (2021), common knowledge "underwrites much of social life". Claims such as these are fairly typical (see also De Freitas et al., 2019). However, these claims rest on a potentially problematic equivocation between common knowledge and public information (Lederman, 2018a). The former is a technical notion defined in terms of higher-order knowledge, while the latter is an informal notion of information that is "completely open" (Heal, 1978) and available to everyone involved.

There is disagreement on whether ordinary people sometimes have common knowledge, but it is not controversial to say that information is sometimes public. For it is not controversial to say that much of our social life and our coordination skills depend on publicly available information, but it does not follow from this, at least not without argument, that our social life and our coordination skills depend on common knowledge. Defenders of common knowledge might therefore accept the following principle.

[1] A different way to explain coordination is by salience (Schelling, 1960; Lewis, 1969; Heal, 1978; Skyrms, 2010). Coordination obtains by salience among creatures that do not engage in epistemic reasoning, such as ants and bacteria (Skyrms, 2010). Perhaps we do so as well (but see Gilbert, 1989). The topic of this paper is coordination "by agreement" (as David Lewis 1969, p. 33, called it), namely by reasoning about common belief or common knowledge.

> If information that p is public in a group, people in the group have common knowledge that p.

As stated, this principle is underspecified. Let us distinguish a claim about ordinary people like us from a claim about the ideally rational people that populate formal models in epistemology.

> COMMON KNOWLEDGE *for ideally rational people*. If information that p is public in a group of ideally rational people, people in the group have common knowledge that p.

> COMMON KNOWLEDGE *for ordinary people*. If information that p is public in a group of ordinary people, people in the group have common knowledge that p.

In this section I will focus on ideally rational people, and I will use COMMON KNOWLEDGE to refer to the version of the principle about ideally rational people, because standard formulations of the arguments I will review below are stated in terms of ideal rationality.

I will assume that ideally rational people conform to standard norms of rationality, such as having consistent preferences and believing what intuitively follows from their beliefs. For specific purposes below, I will assume that ideally rational people update by Bayes rule, and that they prefer to maximize their payoffs. All of these assumptions are familiar. I will not assume that ideally rational people are logically omniscient, nor that they have unlimited memory, energy, or attention (Christensen, 2004; Smithies, 2015; Dogramaci, 2018). I will assume, in other words, only what is necessary in order to predict behavior given the assumptions of a formal model.

There are at least two lines of argument against COMMON KNOWLEDGE: the cognitive objection and the other minds objection. The first is more frequently raised but it is the second that will keep us more occupied. Both objections are meant to suggest that ordinary people don't have common knowledge, despite sharing public information.[2]

2.1 The cognitive objection

According to the cognitive objection, ideally rational people could not have common knowledge, since common knowledge is infinitary and people's minds are finite (Heal, 1978; Parikh, 2005). Since information that p can be public for some p, it follows that COMMON KNOWLEDGE is false. Sperber and Wilson (1995) raise a version of the cognitive objection: if members of a group were to establish that they have common knowledge of something, "they would have, in principle, to perform an infinite series of checks, which clearly cannot be done in the amount of time it takes" to coordinate (p. 18).[3]

A satisfactory reply to the cognitive objection comes in two parts (Paternotte, 2011). The first part is to acknowledge that there is nothing suspicious about finite minds entertaining an infinitary notion. We entertain the notion of a natural number as anything that is either 0 or the successor of a natural number. The second part of the reply is to define common knowledge recursively (Barwise, 1988): something is common knowledge in a group if and only if everyone in the group knows it and knows that it is common knowledge in the group. If members of a group were to

[2]Vanderschraaf and Sillari (2021) provide a more thorough review of debates on the common attitudes. There is a third line of argument against common knowledge, inspired by Williamson's (2000) rejection of the KK principle. Since this objection is not about the link between common knowledge and publicity, I will set it aside—see Greco (2014a, 2023) for discussion.

[3]Sperber and Wilson (1995) use *mutual knowledge* to mean common knowledge, but I will use this term differently below. I follow current usage, as in Fagin et al. (1995), which goes back to Lewis (1969) and Aumann (1976).

establish that something is common knowledge among them, on this definition, they would not have to perform an infinite series of checks. Infinity is not beyond the grasp of finite minds.

Perhaps the challenge raised by the cognitive objection is that of finding something like a basis from which to bootstrap ourselves to common knowledge (Heal, 1978, p. 121). However as Thomason (2021) remarks, the claim that we can never have a justification for common knowledge, as a matter of our cognitive limitations, is not persuasive. Common knowledge, just like individual knowledge, is defeasible and risk-sensitive. Whether someone has knowledge or common knowledge depends on context and background. If a skeptical hypothesis is relevant, I may not be justified in my belief that I have hands, nor in our common belief that we do. Consequently, knowledge and common knowledge may fail. But there seems to be little reason to deny that our beliefs, including our beliefs about each other's beliefs, can be justified if a skeptical hypothesis is not relevant, the risks are low, and circumstances normal. The cognitive objection is not impressive.

2.2 The other minds objection

According to the other minds objection, correct belief-formation is compatible with other minds not thinking alike. Suppose that *a* and *b* are ideally rational people. It is possible for *a* to know that *p*, and to fail to know that *b* knows that *p*: for all *a* knows, *b* might have been distracted, or confused, or might disagree. In these cases, for all *a* knows, *a* and *b* do not share the belief that *p*. Therefore, common knowledge fails. I will discuss three versions of the other minds objection.

Perhaps the better known version of the other minds objection arises from the contingencies of communication. Information can be made public by communication but, contrary to COMMON KNOWLEDGE, it does not thereby amount to common knowledge. Suppose that *a* and *b* are ideally rational people, that *a* knows that *p*, and that *a* sends a message to *b* to share their knowledge. If it is possible that *b* did not receive or did not understand *a*'s message, *a* does not know that *b* knows that *p*, even if *b* does. Of course, *b* could send a message back to confirm, but *b* would not thereby know that *a* knows that *b* knows that *p*, for the same reason. There is no guarantee of common knowledge for any finite number of messages. Communication may be sufficient to make information public, but it is insufficient to establish common knowledge against people's ignorance of other minds.

In a popular illustration of the argument, *a* and *b* are Byzantine generals who must coordinate on attacking their enemy together, but can only do so by sending messengers through enemy lines (Fagin et al., 1995). Since there is no guarantee that any message ever reaches destination, coordination on attacking fails for any finite number of messages. According to Fagin et al. (1999), the lesson of the coordinated attack argument is that "common knowledge cannot be attained when communication is not guaranteed" (p. 92). But communication is hardly ever guaranteed, in a noisy world.

The coordinated attack argument is a challenge about how common knowledge is generated. One reply to the coordinated attack argument is to point out that communication is not the only way to establish common knowledge. A second line of reply is to argue that communication does at least establish that common belief is somewhat probable. Probabilistic common belief approximates common knowledge "from below", as it were, and the two are equivalent for many practical purposes (Monderer and Samet, 1989; Paternotte, 2011, 2017). If probabilistic common belief is attainable, that might be enough for coordination.

However, both replies are ineffective against the second and third version of the other minds objection. The second version is due to Lederman (2018b). Suppose that a and b are ideally rational people who can see some event p and that each other can see it. Since perceptual information, in normal circumstances, is publicly available, a and b have common knowledge that p by COMMON KNOWLEDGE. (In Lederman's example, we assume that two ideally rational people have unobstructed view of a 3m tall sailboat in front of them. In normal circumstances, they have public information that its mast is taller than 1m.) This is one horn of a *reductio* argument. The other horn is an induction, which depends on the assumption that a and b have common knowledge of their mutual ignorance. Let p and p' be distinct events that appear to be arbitrarily similar, neither of which implies the other. Finally, let 'for all x knows, q' be short for 'x does not know that $\neg q$'.

> IDEAL IGNORANCE. It is possible that, if it appears to an ideally rational person that p, then for all they know, it appears to another ideally rational person that p'.

In other words, an ideally rational person's beliefs are compatible with their ignorance of another ideally rational person's beliefs. This seems a reasonable assumption. Suppose that IDEAL IGNORANCE is public information among a and b, and consider the possibility it describes. By COMMON KNOWLEDGE, a and b commonly know that, if it appears to a that p, then for all a knows, it appears to b that p'. (In Lederman's example, a and b commonly know that if it appears to a that the mast is 3m tall, then for all a knows, it appears to b that it is $(3 - \epsilon)$m tall, with ϵ a vanishingly small quantity.) Suppose it appears to a that p. Then a and b commonly know that, for all a knows, it appears to b that p'. Moreover, by the same reasoning, a and b commonly know that, if it appears to b that p', then for all b knows, it appears to a that p''. Thus a and b commonly know that for all a knows, for all b knows, it appears to a that p''. And so on. At the end of the induction, a and b do not have common knowledge that it appears to either of them that p (or anything closely related to p). It follows, plausibly, that a and b do not have common knowledge that p. Lederman concludes that COMMON KNOWLEDGE is false: in some cases, information is public among ideally rational people, and not common knowledge.

Having reached this conclusion in a paradigmatically simple case, Lederman generalizes to other cases, including knowledge based on testimony (communication) or any other kind of justification. Moreover, versions of the same argument can be raised replacing *common knowledge* with *common belief* or *probabilistic common belief*, or any common attitude. Thus the two replies considered above to the first version of the other minds objection are ineffective (Thomason, 2021, p. 237). According to Lederman (2018b), the lesson is that common attitudes such as common knowledge "demand that people be able to access others' minds as if they were their own" (p. 1070). Neither we nor ideally rational people ever have such a privilege.

There are different ways to react to Lederman's argument.[4] The reply I wish to explore is this: the argument is valid, but it is incorrect to blame the contradiction on COMMON KNOWLEDGE. The correct lesson is rather that IDEAL IGNORANCE is false: whenever it appears to an ideally rational person that p, they know that it does not appear to another ideally rational person that p'. Thus, ideally rational people are not ignorant of each other's minds. This conclusion will raise some concerns about how similar we are to the ideally rational people.

[4]There are some possible replies that I will not consider in detail. One is to assume that COMMON KNOWLEDGE is subject to the condition that ideal rationality is public information, and to deny common knowledge of rationality: see Lederman (2017). I set aside this option by running the two cases in parallel: the arguments for ideally rational people on the one hand, and the arguments for ordinary people on the other: see §3 below. Another option is to claim that IDEAL IGNORANCE is never common knowledge. However, we can assume that a public announcement is made reporting to a and b their mutual ignorance. It would seem then that their ignorance is public, and thus common knowledge by COMMON KNOWLEDGE. Similar escape routes for the disagreement argument below are similarly blocked.

Immerman (2021) rejects IDEAL IGNORANCE too, but for a different reason than the one considered here. According to Immerman, IDEAL IGNORANCE is false because sometimes people know that things appear to be a certain way, and moreover, that they could not appear to be some other way. To some extent, Immerman's reply parallels Thomason's anti-skeptical remarks cited earlier: some possibilities are excluded because they are irrelevant, abnormal, or because somehow we know that they do not obtain. Although this may be so, a weakness of Immerman's reply is that it is not general enough to block a third version of the other minds objection, which does not rely on IDEAL IGNORANCE.

The third version of the other minds objection can be extracted from Robert Aumann's (1976) celebrated agreement theorem. The following assumption seems unexceptionable if ideally rational people are people at all.

IDEAL DISAGREEMENT. Ideally rational people can disagree.

Suppose that *a* and *b* are ideally rational people, and that they were once babies. As such, they started off from a sort of "original position" in which they shared a prior. Moreover, suppose that by virtue of their rationality, they learn from experience via Bayes' rule. By IDEAL DISAGREEMENT, consider the possibility that *a* and *b* disagree about *p* for some *p*. Following Aumann, *disagreement* means having different posteriors for *p*. After all, *a* and *b* might have been exposed to different evidence regarding *p*. Finally, suppose that their prior, learning rule, and disagreement, are public information among them. Then, by COMMON KNOWLEDGE, all are common knowledge. By Aumann's famous proof, contradiction follows.

Aumann's argument establishes that if *a* and *b* share prior and update rule and have common knowledge of them, then if they have common knowledge of their posteriors for *p*, they must have the same posterior. So it would be inconsistent to assume that they have common knowledge of different posteriors. As Aumann puts it, ideally rational people "cannot agree to disagree" (p. 1236): they cannot have common knowledge of their disagreement. A skeptic can conclude from this, once again, that COMMON KNOWLEDGE is false.

Aumann's result is, in a sense, quite robust: Geanakoplos and Polemarchakis (1982) and Monderer and Samet (1989) show that analogous results obtain for probabilistic common belief. In another sense, the assumptions of the theorem are quite strong, inviting for possible ways out (Lederman, 2015). For example, Samet (1990) shows that Aumann's theorem depends on a positive introspection axiom on knowledge (if *a* knows that *p* then *a* knows that *a* knows it), which many may find contentious. Another option is to deny that *a* and *b* share a prior: as a reviewer points out, this is a cost-less move for subjective Bayesians. There are other options too.

The three versions of the other minds objection I have discussed are not individually irresistible. Each argument admits of multiple ways out, but many strategies to avoid an unsavory conclusion don't generalize: a retreat to probabilistic common belief helps with the vagaries of communication, but it is of no avail against the arguments from IDEAL IGNORANCE and IDEAL DISAGREEMENT. A rejection of IDEAL IGNORANCE blocks Lederman's argument, but it does not work against the argument from IDEAL DISAGREEMENT. Subjective Bayesians who want to defend common knowledge might be unimpressed by the argument from IDEAL DISAGREEMENT, but have to work out what to say about IDEAL IGNORANCE and the fragility of communication. I conclude from this that there is a general problem raised by the other minds objection that must be reckoned with: common knowledge is incompatible with the observation that ideally rational people may not have a shared understanding of information that is publicly available to them.

The skeptic of common knowledge has a unified account of the general problem raised by the other minds objection: deny COMMON KNOWLEDGE for ideally rational people. This is what Harvey Lederman recommends. Presumably, then, ordinary people too, who are in the same or similar situations as the ideally rational people described by the arguments above, may share public information but no common knowledge. Hence, COMMON KNOWLEDGE for ordinary people is false as well. In the next section, I will deny the purported similarity between us and the ideally rational people. This distinction will create the space for an different unified account of the general problem raised by the other minds objection.

3 Scientific idealizations

The previous section has put into sharp focus two questions that are relevant to understand the consequences of a general strategy to defend common knowledge against various versions of the other minds objection. The first question is whether there are some people who may have common knowledge: COMMON KNOWLEDGE for ordinary people might be true, as Greco (2014b) and Williams (2021) argue, or COMMON KNOWLEDGE for ideally rational people might be true. The second question is about the significance of formal models of collective rational action in which ideally rational people figure prominently. I will take these questions in order.

3.1 Who are the ideally rational people?

Common knowledge has been claimed to be necessary for coordination among ordinary people (De Freitas et al., 2019), and it is certainly necessary for coordination in a wide range of formal models of collective rational action described by Fagin et al. (1995), among others. These models are about ideally rational people. However, in view of the previous section, common knowledge must not be necessary for coordination among ordinary people: if COMMON KNOWLEDGE for ordinary people were true, it would be inconsistent with the following claims by arguments analogous to the ones reviewed in the previous section.[5]

> ORDINARY IGNORANCE. It is possible that, if it appears to an ordinary person that p, then for all they know, it appears to another ordinary person that p'.
> ORDINARY DISAGREEMENT. Ordinary people can disagree.

We are sometimes ignorant of each others' minds and we do sometimes disagree. Therefore, contrary to Greco and Williams, COMMON KNOWLEDGE for ordinary people is false.

If we can't have common knowledge, a weaker principle might explain the effects of public information on ordinary people. Let a and b have *mutual knowledge* that p if and only if both know that p. Public information cannot just be mutual knowledge, because we do reason about what others know. Walking on a busy sidewalk, for example, we try to avoid other pedestrians. Sometimes we look at them, meet their gaze, and guess what they will do. If this happens, coordination may succeed or may not. The awkward "sidewalk dance" we occasionally do, when we try to avoid bumping into others, is evidence that we represent what others know, that they do the same, and that we occasionally make the wrong guess.

[5] As above, the inconsistency would depend on a host of additional and potentially controversial assumptions. One could prevent inconsistency by tiptoeing around some such assumptions, but the aim of the strategy I wish to explore is to provide a unified account of the other minds objection.

Some failures of coordination on the sidewalk can be analyzed as cases in which mutual knowledge is not mutually known: perhaps, both a and b know that a will go left but either a or b does not know that the other knows that a will go left. From here, it is easy to generalize to cases in which people fail to have mutual knowledge at higher iterations of the knowledge operator, resulting in more and more sophisticated cases of coordination failure (Greco, 2015; Lederman, 2018a). Let a and b have *mutual knowledge*$^{n+1}$ that p if and only if they both know that they have mutual knowledge^n that p, and let mutual knowledge^0 be mutual knowledge *simpliciter*. Lederman (2018b, p. 1095, fn. 22) conjectures that a version of the following principle, for a relatively small n, is a sufficient characterization of the effects of public information on ordinary people for a variety of scientific applications.

> MUTUAL KNOWLEDGE^n. If information that p is public in a group of ordinary people, people in the group have mutual knowledge^n that p.

At infinity, mutual knowledge$^\omega$ is just common knowledge. Of course, we can't iterate all the way to ω, on pain of inconsistency with ORDINARY IGNORANCE and ORDINARY DISAGREEMENT. However, if all we have is finitely iterated mutual knowledge, often called *almost-common knowledge* or *truncated common knowledge*, then the contradictions of the previous section can be explained away (Lederman, 2018b, 1091).

Suppose that a and b are ordinary people who can see some event p and that each other can see it, as in Lederman's scenario. By MUTUAL KNOWLEDGE^1, ORDINARY IGNORANCE is false. Since a knows that p, it appears to a that p. Moreover a knows that it does not appear to b that p', for a knows instead that b knows that p. In addition, since a and b have mutual knowledge, they do not disagree—at least not about p. Then ORDINARY DISAGREEMENT about p is false as well. Consistency is upheld whenever some people have mutual knowledge^n that p, if p is public information, but not mutual knowledge^m of ORDINARY IGNORANCE and ORDINARY DISAGREEMENT about p, for $m \geq n$. According to this line of reasoning, ordinary people get "off the hook", so to speak, because they have no more than almost-common knowledge.

There is still a question about the ideally rational people: are the skeptical arguments of the previous section sound for them too? Although Lederman suggests that ideally rational people don't have common knowledge, an intriguing possibility is that, while COMMON KNOWLEDGE for ordinary people is false, COMMON KNOWLEDGE for ideally rational people is true. If so, models of collective rational action in economics, linguistics, and social science, are not threatened by the arguments of the previous section. After all, it is the ideally rational people who figure prominently in formal models of rational action and who coordinate by reasoning about common knowledge.

However, if public information is common knowledge for ideally rational people, then ideally rational people are very different from us. Unlike us, they do not have disagreements and they are not ignorant of each others' minds. By pursuing this strategy, we have a unified account of the general problem raised by the other minds objection, but a further question becomes relevant. The assumption of common knowledge generates significant distortions between models of collective rational action and the empirical reality they are supposed to be models of: what is the relation between us and the ideally rational people?[6]

[6]Another interesting question that arises at this point is about the interpretation of formal models based on common knowledge and related notions, such as Stalnaker's (2012) influential model of conversation. See Daniel Harris (2020) for a discussion of this point.

3.2 Discontinuity at infinity

A seemingly widely shared view is that ideally rational people are abstractions or simplifications of ordinary people: the effects of public information on our collective rational actions are finite approximations of the effects of common knowledge on ideally rational people. Along these lines, Williamson (2000) claims that common knowledge is "a convenient idealization, like a frictionless plane" (p. 122). Similarly, Lederman (2017) argues that common knowledge is "just a simplifying technical assumption, which is useful because it yields tractable models and rich predictions about behavior" (p. 15). Simplifying technical assumptions in science are known as *Galilean idealizations* (Frigg and Hartmann, 2020).

In a Galilean idealization, there is a trade-off between the truth of a hypothesis and its computational complexity, which is resolved by keeping the assumptions of the model approximately true and the mathematics relatively simple. The analogy, following Williamson, is with the frictionless plane: we may calculate a body's velocity while neglecting friction, since adding its effects does not change our understanding of the phenomena and comes at the price of a considerably more complicated formalism. In a Galilean idealization, the price of empirical accuracy is the loss of mathematical simplicity. Importantly for present purposes, the distortions introduced by a Galilean idealization, despite being actual falsehoods (Jones, 2005), do not undermine the approximate truth of the theory. The assumption that there is no friction is false, but the theory of motion described under that assumption is approximately true.

If common knowledge is a Galilean idealization, more realistic models of collective rational action are obtained by assuming almost-common knowledge, just like more realistic models of motion are obtained by factoring in the effects of non-zero friction. Both common knowledge and the frictionless plane can be understood as infinite limits. Common knowledge may be regarded as the limit for $n \to +\infty$ of mutual knowledge^n. Friction can be understood in terms of a coefficient $1/x$ for $x \to +\infty$ so that, at infinity, the contribution of friction is null. An important consequence of the trade-off implicit in a Galilean idealization is about *emergence*, or rather, lack thereof:[7] in a Galilean idealization, the idealized model is reducible, in some sense, to the more realistic and less idealized model (McMullin, 1985).

Physicists are entitled to the study of approximate physics, since the predictions of a theory that assumes non-zero friction approximate the predictions of idealized theories that assume no friction, as $1/x$ goes to zero. Table surfaces are finitary approximations of frictionless planes. On the Williamson/Lederman view, ordinary people are finitary approximations of ideally rational people. Accordingly, economists, linguists, and social scientists are entitled to the study of common knowledge as well—provided the same kind of similarity holds between the finitary but mathematically complicated model that describes us, and the idealized infinitary model. Therefore, modulo the proviso I just mentioned, the Williamson/Lederman view that common knowledge is a Galilean idealization does not force a revision of scientific practice.[8]

However, not all infinite limits are Galilean idealizations (Callender, 2001; Butterfield, 2010; Batterman, 2011). In particular, if common knowledge were a Galilean idealization, then the behavior of rational people in the finitary theory, which only assumes almost-common knowledge,

[7] I take emergence and reduction to be mutually exclusive: theory T_1 is emergent given a base theory T_1 iff T_1 is not reducible to T_2. There are several different notions of emergence, hence of reduction: see Palacios (2022) for an overview. Below, I will take a stand on what we might mean by *emergence*, and thus *reduction*, in the present context.

[8] For a critical discussion of the Williamson/Lederman view, based on different considerations than the ones mentioned here, see Greco (2023, pp. 164ff). The lesson I will draw below about the value of common knowledge in scientific practice cannot be extracted from, and might even be incompatible with, Greco's discussion.

would approximate the behavior of ideally rational people in the infinitary theory, which assumes common knowledge. But it does not. Hence, the analogy between ideally rational people and the frictionless plane is misguided. Common knowledge is not a Galilean idealization since it is not approximated by mutual knowledge^n for finite n in some relevant cases: sometimes, the predictions of collective rational action models at infinity are not close to the predictions of such models at any point below infinity, and they do not get any closer no matter how much mathematical complexity we factor in (within a finite theory). There is discontinuity at infinity.

An illustration of this point is Ariel Rubinstein's (1992) electronic mail game. Consider two ideally rational people a and b, who play game G_p if p and game $G_{\neg p}$ if $\neg p$, with payoffs as illustrated in Table 1. The game models a situation in which, depending on whether p or $\neg p$, two partners profit from coordinating on different actions: A in the first case, B in the second.

	G_p			$G_{\neg p}$	
	A	B		A	B
A	1,1	0,-2	A	0,0	0,-2
B	-3,0	0,0	B	-3,0	2,2

Table 1: Electronic Mail Game

Suppose that p is more likely than $\neg p$. If p, then a and b coordinate on A as doing otherwise would be irrational by standard rational choice. If $\neg p$ then they coordinate on B, for similar reasons. Suppose however that only a knows whether p or $\neg p$, and if the latter then a sends an e-mail to b to make sure b knows that $\neg p$, whereas if p then no e-mail is sent. There is also a small but non-zero probability that an e-mail is not delivered due to a glitch in the system. So, upon receiving a's e-mail, b ought to make sure that a knows that b knows, and thus a confirmation e-mail is sent back. And so on. If a and b exchange infinitely many e-mails then they achieve common knowledge that $\neg p$, hence they coordinate on B. But if they only exchange finitely many e-mails, they do not have common knowledge that the game is $G_{\neg p}$. In this case, if a chooses B then a's expected payoff is less than it would be if they chose A, since p is more likely than not, and the possibility remains that b might choose A. Hence a chooses A, and so does B for the same reasons. Therefore, if a and b have common knowledge that $\neg p$ then they both choose B, otherwise they both choose A. In this case, rational behavior at the limit is qualitatively different from rational behavior at any point below the limit.[9]

A discontinuity is a distortion that is not an approximation. Any Galilean idealization introduces some distortion. For example, an object traveling on a frictionless plane is predicted to move forever. Objects clearly do not move forever, but it is approximately true that they do, in the following sense: as the friction coefficient gets smaller and smaller, the object's travel gets longer and longer. The behavior of the system at the limit is thus continuous with the behavior of the system below the limit, as the value of the relevant variable approaches the limit. Eventually, and with sufficiently complex mathematics, the predictions of the finitary theory match those of the infinitary theory. In this sense, a Galilean idealization is a distortion, but also an approximation. Common knowledge is different. It is a distortion but not an approximation. In the electronic mail game, it is not true that for greater and greater iterations of mutual knowledge^n, a and b get closer

[9] Rubinstein's (1992) result depends on a few unrealistic assumptions. In particular, the e-mails are sent automatically and are costless. As shown by Binmore and Samuelson (2001), if the e-mail exchange is strategic or costly, there are equilibria in which the players coordinate on B after finitely many messages. However, even though Rubinstein's result holds under unrealistic conditions, what matters is the discontinuity it reveals between finitary and infinitary models.

and closer to coordinating on B: no matter how large (but finite) is n, a and b coordinate on A. They simply don't get any closer to coordinating on B as n goes to infinity. So there is no trade-off between empirical accuracy and mathematical complexity, because complicating the mathematics (within a finitary theory) does not change the outcome.

The electronic mail game shows that there can be discontinuity at infinity: the behavior of ideally rational people who share common knowledge need not be approximately similar to the behavior of people who share at most almost-common knowledge. Infinite limits at which the predictions of the theory are discontinuous with respect to its predictions at any point below the limit are not Galilean idealizations. The analogy between ideally rational people and the frictionless plane breaks down. So, common knowledge is not merely a convenient mathematical simplification. We need a different interpretation of the scientific and explanatory value of common knowledge, and of the ideally rational people who can share it.

3.3 How-actually and how-possibly

I suggest to take the discontinuity at the limit to be evidence of an emergent phenomenon—and I should flag immediately that "emergence" is a controversial term (Fletcher et al., 2019), to be clarified below. Intuitively, in an emergent phenomenon, the system at the limit has fewer "degrees of freedom" than the system at any point approaching the limit. The paradigmatic case of an emergent phenomenon in physics is a phase transition: for example, the sudden transformation of a liquid into gas, depending on temperature and pressure, such as the vaporization of water in a teapot (Humphreys, 2016). In this case, a finitary model of the phenomenon supports true premises, such as, in particular, a premise stating that the number of water molecules in the teapot is finite. In an infinitary model, in contrast, we assume that the number of molecules in the teapot is infinite: the assumption is false, but it leads to a description of the phenomenon under investigation that is discontinuous with the predictions delivered by the finitary model (and, in addition, to a description that is empirically accurate). In this example, theories in which we assume that the number of molecules in the teapot is greater and greater (but still finite) do not approximate the predictions of a theory in which we assume infinitely many molecules.

Discontinuity is evidence of emergence, in the sense I will use this word. The reason we observe a gap between the predictions of a false infinitary assumption and those of a true finitary one is that the phenomenon under investigation does not depend on the variability left open by a finitary model: to account for phase transitions we need to know about temperature and pressure, not about the number of molecules. Assuming infinitely many molecules is a way of bracketing questions about how many molecules are necessary to account for the phenomenon.

The case of common knowledge is similar. In the study of phase transitions, we are not interested in the quantity of water molecules in a teapot. Likewise, in the study of coordination, we are not interested in some of the variation ("degrees of freedom") left open by almost-common knowledge: we need to know about several rationality assumptions verified by the model, but not about how many iterations of the mutual knowledge operator the agents in the model satisfy. Increasing iterations of mutual knowledge need not lead the finitary and the infinitary theory to similar predictions.

Thus, coordination is not to be explained by reference to ignorance, communication, disagreement, or lack thereof. Common knowledge forecloses some possibilities left open by almost-common knowledge that are irrelevant for an account of the phenomena under investigation. In this sense, coordination is an emergent phenomenon. Analogously, phase transitions are emergent: the transition of water into vapor at high temperature and pressure is not to be explained by reference to the number of molecules of water in the teapot. The thermodynamic limit is a better analogy to common knowledge than the frictionless plane.

The important distinction between models analogous to the frictionless plane and models analogous to a teapot containing infinitely many molecules is not about empirical accuracy.[10] On the view I am suggesting, it is the discontinuity between common knowledge models and mutual knowledge^n models, as illustrated in the electronic mail game, that shows that common knowledge models are not just mathematical simplifications of more realistic models. There is no such discontinuity between frictionless models of motion and more realistic models that account for friction: neglecting friction simplifies the math but does not change our understanding of the phenomena. Ideally rational people may be simpler versions of us but they are also qualitatively different from us.

As I mentioned, "emergence" has multiple interpretations (O'Connor, 2021). I will conclude this paper by discussing at least two versions of the claim that coordination is an emergent phenomenon, to offer an account of why it might still be interesting to study the behavior of ideally rational people, despite their not being similar to us. The two versions of emergence I'll discuss depend on an important distinction between how-actually and how-possibly questions (Reutlinger et al., 2018).

One view is that common knowledge is the answer to the question: how do ideally rational people actually coordinate? Such a *how-actually* interpretation of common knowledge models appears to tie in with a metaphysical interpretation of emergence. Since common knowledge is not something ordinary people can have, ideally rational people are metaphysically on a different level of reality than us, and our study of coordination by means of models that assume common knowledge are studies into their reality, not ours.

The how-actually interpretation might be supported by the claim that formal models of collective rationality are not about empirical reality, but rather a form of sophisticated storytelling (Sugden, 2000). According to this view, models of collective rationality describe counterfactual scenarios whose purpose is to allow us to make inductive inferences that may be of practical significance. The claim of emergence implies that such counterfactual worlds are not metaphysically reducible to our world. Alternatively, the how-actually interpretation might be supported by an anti-realist view of scientific modeling, in particular fictionalist views (Contessa, 2010; Frigg, 2010;

[10] Philosophers of physics disagree on whether the infinitary limit is "essential" to the phenomenon of phase transitions (Callender, 2001; Batterman, 2011; Butterfield, 2010). This is why *emergence* is controversial in this context. Those who think that the discontinuity is inessential prefer to avoid the word "emergence" and claim that a better approximation of empirical reality is obtained by "simulating" limit behavior by means of very skewed (but still continuous) sigmoid functions. This debate depends on comparing the predictions of a discontinuous model with those of an almost-discontinuous one. There is an analogous question about collective rationality (Greco, 2023, p. 196): is the empirically observable behavior of ordinary people in a group better represented by models that assume common knowledge, or by models that assume almost-common knowledge and somehow "simulate" common knowledge in a finitary setting? I think we don't know, and for the purposes of my argument this question may be left open. Thus, I will continue to use the word "emergence". It is somewhat of a thorny issue whether common knowledge is necessary for empirically accurate predictions in economics and other disciplines, and more generally whether formal models of collective rationality are (ever?) empirically accurate. See Hausman (1984), Sugden (2000), Mäki (2012), and Hausman (2018) for an overview. Regarding experimental simulations of the electronic mail game, results are not decisive (Kawagoe and Takizawa, 2012). More generally, research that compares models that assume less than common knowledge with empirical findings has led to mixed results. For an early discussion, see Bicchieri (1993). For more recent work, see Binmore (2008), Crawford et al. (2013), and references therein. It is a strength of my argument, at least given the present state of our knowledge, that it does not depend on comparing the accuracy of the predictions of different models of collective rationality. I would like to thank an anonymous reviewer for prompting me to clarify this issue.

Levy, 2015). Although both counterfactual and fictionalist accounts of common knowledge are consistent with my conclusions, they are not supported by them without further arguments.

On a different interpretation, common knowledge is useful (though maybe not necessary) to answer the question: how could ideally rational people coordinate? This *how-possibly* interpretation ties in nicely with an epistemic interpretation of "emergence". The conceptual repertoire of common knowledge delivers predictions about rational behavior that cannot be explained or derived from assumptions about almost-common knowledge. An account of coordination in terms of common knowledge holds at a level of explanation not accessible to a finitary theory.

The how-possibly interpretation yields a straightforward account of impossibility theorems, such as the results of Aumann (1976), Rubinstein (1992), and Lederman (2018b), that we mentioned above. Herein certainly lies some of the interest for common knowledge. A natural way to ask a how-possibly question is to ask how something is not impossible: how is coordination not impossible, among ideally rational people? Impossibility results such as the ones above are descriptions of scenarios on which coordination is impossible. Conversely, coordination can be explained if at least some of the assumptions of these theorems fail. There is no reason to draw a further conclusion from here, however, concerning how ideally rational people actually coordinate. Even less reason to draw a conclusion concerning how we ordinary people actually coordinate.

Still, it may be important to answer how-possibly questions. The how-possibly interpretation of common knowledge models dovetails nicely with David Lewis's (1969) influential work on conventions, from which much philosophical discussion on common knowledge comes. Lewis's work was a response to Quine (1936), who had argued that language cannot be conventional, since conventions must be agreed upon by all parties, and language is needed to reach an agreement (Lewis, 1969, p. 2). Lewis showed that conventions can be established without language, by assuming something like common knowledge (though not the modern notion of common knowledge: see Cubitt and Sugden, 2003, and Sillari, 2008). Importantly, although Lewis showed that common knowledge is sufficient to coordinate on a convention, he never said that it is necessary: he showed that there can be "tacit agreement" via common knowledge, besides the explicit agreement that depends on sharing a language. So language can be conventional after all, despite Quine. A useful appeal to common knowledge, interpreted how-possibly, has been to establish the possibility of linguistic conventions. No conclusion follows from this concerning how ideally rational people actually coordinate on which language to use, even less concerning us.

4 Conclusion

I have discussed skeptical arguments against common knowledge, and focused on three versions of the other minds objection. I have defended common knowledge, but I have done so by accepting the skeptical conclusions. The skeptic is right to say that common knowledge is unattainable for us. Nevertheless, we ought not reject models of coordination that appeal to it. A good reason for this is that common knowledge might be useful to explain how certain social configurations (such as a linguistic convention) are not impossible. Such an explanation might require conceptual resources that are unavailable at the level of almost-common knowledge.

I have argued that common knowledge models may be taken to describe the behavior of ideally rational people, provided such idealizations are not understood as approximating anything in the empirical world—in particular, they are not merely mathematical simplifications of ordinary people. It follows that a straightforward connection between common knowledge models and their targets, analogous to the frictionless plane, is flawed. On a more appropriate way of understanding the theoretical role of ideally rational people, common knowledge is an infinite limit that is necessary to describe an emergent phenomenon: rational coordination in a group of people cannot be reduced, in all cases, to the epistemic profiles of individual members of the group, no matter how much they (finitely) know that others know that they know. There are different ways to interpret this claim about emergence, or failure of reduction, and I have sketched some reasons why at least an epistemic interpretation is plausible.

References

Aumann, R. (1976). Agreeing to disagree. *The Annals of Statistics 4*, 1236–1239.

Barwise, J. (1988). Three views of common knowledge. In M. Vardi (Ed.), *Proceedings of the 2nd conference on theoretical aspects of reasoning about knowledge*, pp. 365–379. San Francisco, CA: Morgan Kaufmann Publishers Inc.

Batterman, R. W. (2011). Emergence, singularities, and symmetry breaking. *Foundations of Physics 41*, 1031–1050.

Bicchieri, C. (1993). *Rationality and Coordination*. New York, NY: Cambridge University Press.

Binmore, K. (2008). Do conventions need to be common knowledge? *Topoi 27*(1-2), 17–27.

Binmore, K. and L. Samuelson (2001). Coordinated action in the electronic mail game. *Games and Economic Behavior 35*, 6–30.

Butterfield, J. (2010). Less is different: Emergence and reduction reconciled. *Foundations of Physics 41*(6), 1065–1135.

Callender, C. (2001). Taking thermodynamics too seriously. *Studies in History and Philosophy of Science 32*, 539–553.

Christensen, D. (2004). *Putting Logic in its Place*. Oxford: Oxford University Press.

Contessa, G. (2010). Scientific models and fictional objects. *Synthese 172*, 215–229.

Crawford, V. P., M. A. Costa-Gomes, and N. Iriberri (2013). Structural models of nonequilibrium strategic thinking: Theory, evidence, and applications. *Journal of Economic Literature 51*(1), 5–62.

Cubitt, R. and R. Sugden (2003). Common Knowledge, Salience, and Convention: A Reconstruction of David Lewis's Game Theory. *Economics and Philosophy 19*, 175–210.

De Freitas, J., K. Thomas, P. DeScioli, and S. Pinker (2019). Common knowledge, coordination, and strategic mentalizing in human life. *Proceedings of the National Academy of Sciences 116*(28), 13751–13758.

Dogramaci, S. (2018). Solving the problem of logical omniscience. *Philosophical Issues 28*(1), 107–128.

Fagin, R., J. Y. Halpern, Y. Moses, and M. Vardi (1995). *Reasoning About Knowledge*. MIT Press.

Fagin, R., J. Y. Halpern, Y. Moses, and M. Y. Vardi (1999). Common knowledge revisited. *Annals of Pure and Applied Logic 96*, 89–105.

Fletcher, S. C., P. Palacios, L. Ruetsche, and E. Shech (2019). Infinite idealizations in science: an introduction. *Synthese 196*, 1657–1669.

Frigg, R. (2010). Models and fiction. *Synthese 172*, 251–268.

Frigg, R. and S. Hartmann (2020). Models in Science. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Spring 2020 ed.). Metaphysics Research Lab, Stanford University.

Geanakoplos, J. D. and H. M. Polemarchakis (1982). We can't disagree forever. *Journal of Economic Theory 28*, 192–200.

Gilbert, M. (1989). Rationality and salience. *Philosophical Studies 57*, 61–77.

Greco, D. (2014a). Could KK be OK? *The Journal of Philosophy 111*, 169–197.

Greco, D. (2014b). Iteration and fragmentation. *Philosophy and Phenomenological Research 88*, 656–673.

Greco, D. (2015). Iteration principles in epistemology i: Arguments for. *Philosophy Compass 10*(11), 754–764.

Greco, D. (2023). *Idealization in Epistemology*. Oxford: Oxford University Press.

Harris, D. W. (2020). We talk to people, not contexts. *Philosophical Studies 177*, 2713–2733.

Hausman, D. M. (1984). *The Philosophy of Economics: An Anthology*. New York: Cambridge University Press.

Hausman, D. M. (2018). Philosophy of economics. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Fall 2018 ed.). Metaphysics Research Lab, Stanford University.

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

Humphreys, P. (2016). *Emergence: A Philosophical Account*. Oxford: Oxford University Press.

Immerman, D. (2021). How common knowledge is possible. *Mind forthcoming*, 1–15.

Jones, M. R. (2005). Idealization and abstraction: A framework. *Poznan Studies in the Philosophy of the Sciences and the Humanities 86*(1), 173–218.

Kawagoe, T. and H. Takizawa (2012). An experimental study of e-mail games with strategic information transmission and communication cost. *Economics Bulletin 32*(4), 2921–2929.

Lederman, H. (2015). People with common priors can agree to disagree. *Review of Symbolic Logic 8*(1), 11–45.

Lederman, H. (2017). Two paradoxes of common knowledge: Coordinated attack and electronic mail. *Noûs 52*, 921–945.

Lederman, H. (2018a). Common knowledge. In M. Jankovic and K. Ludwig (Eds.), *The Routledge Handbook of Collective Intentionality*, pp. 181–195. New York: Routledge.

Lederman, H. (2018b). Uncommon knowledge. *Mind 127*(508), 1069–1105.

Levy, A. (2015). Modeling without models. *Philosophical Studies 152*, 781–798.

Lewis, D. (1969). *Convention*. Cambridge MA: Harvard University Press.

Mäki, U. (2012). Realism and antirealism about economics. In U. Mäki (Ed.), *Handbook of the Philosophy of Economics*, pp. 3–24. Amsterdam: Elsevier.

McMullin, E. (1985). Galilean idealization. *Studies in History and Philosophy of Science Part A 16*(3), 247.

Meyer, J.-J. C. and W. van der Hoek (1995). *Epistemic Logic for Computer Science and Artificial Intelligence*. Cambridge: Cambridge University Press.

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

Osborne, M. J. and A. Rubinstein (1994). *A Course in Game Theory*. Cambridge, MA: MIT Press.

O'Connor, T. (2021). Emergent Properties. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2021 ed.). Metaphysics Research Lab, Stanford University.

Palacios, P. (2022). *Emergence and Reduction in Physics*. Cambridge: Cambridge University Press.

Parikh, R. (2005). Logical omniscience and common knowledge: what do we know and what do we know? In R. van der Meyden (Ed.), *Proceedings of the 10th Conference on Theoretical Aspects of Rationality and Knowledge*, pp. 62–77. Singapore: National University of Singapore.

Paternotte, C. (2011). Being realistic about common knowledge: A lewisian approach. *Synthese 183*, 249–276.

Paternotte, C. (2017). The fragility of common knowledge. *Erkenntnis 82*(3), 451–472.

Quine, W. V. (1936). Truth by convention. *Journal of Symbolic Logic 1*, 77–106.

Reutlinger, A., D. Hangleiter, and S. Hartmann (2018). Understanding (with) toy models. *British Journal for the Philosophy of Science 69*(4), 1069–1099.

Rubinstein, A. (1992). The electronic mail game: Strategic behavior under "almost common knowledge". In *Knowledge, Belief, and Strategic Interaction*, pp. 317–326. Cambridge University Press.

Samet, D. (1990). Ignoring ignorance and agreeing to disagree. *Journal of Economic Theory 52*, 190–207.

Schelling, T. C. (1960). *The Strategy of Conflict*. Cambridge, MA: Harvard University Press.

Schiffer, S. (1972). *Meaning*. Oxford: Clarendon Press.

Sillari, G. (2008). Common knowledge and convention. *Topoi 27*(1-2), 29–39.

Skyrms, B. (2010). *Signals: Evolution, Learning, and Information*. New York, NY: Oxford University Press.

Smithies, D. (2015). Ideal rationality and logical omniscience. *Synthese 192*, 2769–2793.

Sperber, D. and D. Wilson (1995). *Relevance: Communication and Cognition*. Oxford: Blackwell.

Stalnaker, R. (2002). Common ground. *Linguistics and Philosophy 25*(5–6), 701–721.

Stalnaker, R. (2012). *Mere Possibilities: Metaphysical Foundations of Modal Semantics*. Princeton, NJ: Princeton University Press.

Stalnaker, R. (2014). *Context*. Oxford University Press.

Sugden, R. (2000). Credible worlds: the status of theoretical models in economics. *Journal of Economic Methodology 7*, 1–31.

Thomason, R. H. (2021). Common knowledge, common attitudes, and social reasoning. *Bulletin of the Section of Logic 50*(2), 229–247.

Vanderschraaf, P. and G. Sillari (2021). Common knowledge. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Fall 2021 ed.). Metaphysics Research Lab, Stanford University.

Williams, J. R. G. (2021). Publicity and common commitment to believe. *Erkenntnis*, 1–22.

Williamson, T. (2000). *Knowledge and its Limits*. Oxford University Press.