

# Paradoxes of Common Knowledge Revisited: A Perspective from Game Theory and Economics

## Abstract

Stephen Morris  
Department of Economics  
University of Pennsylvania  
3718 Locust Walk, Philadelphia PA 19104  
Email: smorris@econ.sas.upenn.edu

Common knowledge and thus perfect coordination are impossible in any information system where there are communication errors. This is true however unlikely the communication errors and however long and detailed the communication. This is the canonical example of a "paradox" of common knowledge; such paradoxes have been studied in recent years by researchers in computer science, philosophy, economics and other fields. The study of such paradoxes served two important functions. First, researchers attempting to account for them developed formal ways of thinking about interactive knowledge. Second, because the same paradoxes attracted attention across many fields, they forced (encouraged?!) researchers in the different fields to talk to each other.

But granted that these two functions were successfully performed, do we have anything more to learn from the paradoxes? In practical problems, something less than common knowledge ("almost common knowledge") is often enough to achieve co-ordination. Is the extreme case of common knowledge instructive about problems of co-ordination or is an intellectual curiosity with few implications for practical problems of co-ordination? These questions are the subject of this discussion session. I shall answer them with a qualified yes.

Before doing so, I want to note how an economist's perspective on these problems will differ from others' perspectives because of strategic issues. First, a physical description of a communication protocol is not sufficient for the economist. It is important to also check that communicators have an incentive to tell the truth. Second, even granted honest communication, the relevant notion of almost common knowledge is going to depend on the kind of co-ordination problem studied, and thus the field. Let me illustrate this with the co-ordinated attack problem. Consider the version where a first general sends an instruction to attack to a second; confirmations are sent back and forth until one gets lost (each message is lost with some small probability), and the two generals must attack simultaneously at dawn. It

is impossible to describe a rule such that (1) attack occurs with positive probability and (2) it is never the case that one general attacks alone. The economist, on the other hand, would typically be perfectly happy to ensure that it is exceedingly unlikely that one general attacks alone. A simple rule which ensures this is for the first general to always attack and the second general to attack if he has received at least one message. But can the generals be relied on to follow such an instruction? Suppose that the generals do not necessarily follow orders, and that each general's preferences are such that he would not want to attack unless he attached probability significantly greater than half to the other general attacking. Then this simple rule would not work. If the first general had received no confirmation, he would attach probability about one half to his first message never having arrived. Since the second general would not attack in that contingency, the first general will not attack. A similar argument by contradiction can be used to show that no attack ever happens in any equilibrium of this incomplete information game. Thus even if the analyst is prepared to accept a small probability of failure of co-ordination, this small probability may become large because of strategic concerns. Since processors presumably do not have different preferences, this issue does not arise in computer science.

Subject to these two provisos, economists should join with other researchers in answering the following questions:

1. Given that perfect co-ordination is not possible with imperfect communication, what is the best that can be done subject to a given information system / communication protocol?
2. To the extent that information systems / communication protocols can be designed (and - for the economist - made incentive compatible), how should they be designed?

I will report on some work in economics addressing these questions and make suggestions for future work.