What Evolutionary Biology Can Tell Us About Cooperation (and Trust) in Online Networks

Toby Handfield, Monash University, toby.handfield@monash.edu

—————————————

In their introduction to this special issue, Alfano and Klein (2019) pose two neatly contrasting questions for social epistemologists who want to take our epistemic networks seriously. First, what sort of *individual* epistemic properties should we cultivate, given the social networks in which we reside; and second, what sort of *networks* should we cultivate, given our individual epistemic dispositions. These are excellent questions, but separating the two questions in this fashion may create an impression that the two issues can be separated, not merely in theory, but in practice. Drawing on considerations from the study of cooperation in evolutionary biology, I will suggest that many online epistemic behaviors serve functions that are also relevant to the maintenance and constitution of social networks. Consequently, it is difficult to envisage any way in which either individual behaviors or network structures could be modified in isolation. An adequate social epistemology needs to answer Alfano and Klein's two questions, not one at a time, but simultaneously.

Cooperation Versus Trust

Trust is a pre-requisite of cooperation, and cooperation is often valuable. This is a central reason why trust is of interest. But trust has its darker side (Baier 1986): trust opens us up to exploitation, and trust can facilitate cooperation among malefactors. I may trust my blackmailer not to publish what he knows about me if I make the demanded payment: but this is not a sort of trust that makes me better off. In fact, if I did not trust him in this way, then he would not be able to blackmail me at all. In such cases trust can be downright undesirable. For this reason, I suggest it is more helpful—for my present purposes at least—to focus on cooperation. Cooperation, understood as two or more individuals coordinating their behavior so as to achieve a larger collective good than would otherwise occur—is always (*pro tanto*) good. Of course cooperation between negligent or malevolent parties may generate negative externalities for others, making the cooperation less than good, all things considered. But we can reasonably bracket those costs precisely because they are *external* to the cooperative relationship. Whereas, in at least some of the cases Baier warns us about, the ills associated with trust are internal to the relationship itself, not merely side effects of that relationship.

My question, then, is what lessons might epistemologists draw about the possibilities for cooperation in online networks, from evolutionary biology and evolutionary game theory in particular. If we can identify where it is possible to achieve cooperation, we are almost invariably also identifying possibilities for sustaining trust. And conversely, where cooperation is impossible, this will frequently be because of the inability to generate trust in that environment.

The conclusions I obtain shall be conditional. *If* certain analogies hold between evolutionary dynamics and the dynamics of individual behavior in these settings, then we may expect cooperation to be fostered in particular conditions. While I will make the relevant analogy clearer below, and sketch the methodological impetus for thinking the analogy has some merit, I won't in this piece attempt to establish *whether* the relevant analogy holds, but leave that for future work.

Cooperation in Online Settings

It is a touchstone of social epistemology that our epistemic achievements are irredeemably dependent on evidence that is obtained via chains of testimony. The network of inquirers in which we are embedded is what delivers that testimony, and hence is crucial to my ability to know things and get things done. I know, for instance, that arithmetic is either incomplete or incoherent: it contains unprovable truths or it contains contradictions. I have no ability to prove this myself, but I have learned from teachers, who themselves learned from books or other teachers, about Gödel's ingenious creation of a sentence of formal mathematics that effectively asserts *of itself* that it is true but unprovable. The technique involved in constructing that sentence is formidable—probably fewer than 1% of the population could employ this technique, even with significant training—yet many more than 1% of the population is in a position to know what follows from constructing that sentence.

The activities of such epistemic networks—when things go well—achieve more knowledge (both collective and individual) than would be achieved if the network did not exist. If my epistemic network had been severed before I took a formal logic class as an undergraduate— no teachers, no books, no peers—I never would have known about the incompleteness of arithmetic. The functioning of the network requires a sort of coordinated activity: events of testimony, challenging of that testimony, elaboration, recapitulation, and so on. And the network achieves a collective (epistemic) good that would not otherwise be available. So this is a sort of cooperation, as I understand it. Unsurprisingly, some degree of trust is required for this cooperation—and also unsurprisingly, *unqualified* trust is inimical to achieving this good. If we gullibly accepted all the testimony we received about mathematics, we would by now believe a number of falsehoods, and probably several contradictions also. All manner of mathematical cranks would be able to pass off their falsehoods upon us.

These examples, however, are rather exalted and academic, and may not much resemble the sort of epistemic activity that most of us think of when we think of *online* networks—the focus of this special issue. Twitter users, for instance, have spent a good deal more time and effort sharing testimony regarding Obama's alleged birthplace, or regarding Putin's alleged meddling with US powergrids, than they have spent on exchanging information about the incompleteness of arithmetic. Is it similarly helpful to model these interactions as a form of cooperation?

Relatedly: if behavior on Twitter, Reddit, Facebook and the like are types of cooperation, what is the relevant "collective good" that is generated? Is said good something relatively objective, such as *knowledge*, or is it relatively subjective, such as the satisfaction of individual desires?

Those with epistemic concerns would naturally adopt a framing in which cooperation requires the advancement of knowledge. And consequently, online networks that spread far-fetched rumours and conspiracy theories are candidates to be *failures* of cooperation. The collective may be epistemically worse off, as a result of their activity, than they would otherwise have been. From a more pragmatic perspective, however, it may be more useful to

adopt a subjective understanding of "the good", akin to that used in rational choice theory. On such an understanding, whenever agents coordinate to achieve a Pareto superior outcome—an outcome preferred by at least one person, and not dispreferred by any—they have successfully cooperated. (Again, we can ignore externalities as a simplifying device: only the preferences of those involved in the online network should be considered.)

There is a methodological reason to prefer this more subjective understanding of cooperation, at least for the present project: the relevant tools of evolutionary biology make predictions about what behaviors or traits will become prevalent in the population by representing outcomes as having impacts on fitness, where fitness is functionally defined such that more fit types become more numerous in the next generation. In game theory, there is a direct structural analogy between *utility* and *fitness*, so understood. Whatever utility is, for game theory to be useful, options that are associated with higher utility must be the ones that agents will reliably choose. Since we don't have any reason to think agents will reliably choose true beliefs rather than false ones, but we do have reason to think that they will choose preferred options over dispreferred ones, we have strong reason to think that utility must track subjective states like preferences. It is because of this analogy between fitness and utility that evolutionary biologists have been able so successfully to adapt the techniques of game theory to biological settings (Maynard Smith 1974).

So with this non-judgmental understanding of cooperation, we are in a better position to draw conclusions from biology about when cooperation will arise, but we are consequently in a less strong position to decide whether or not that cooperation will be epistemically virtuous or vicious. A group of conspiracy theorising twitter-users, on this understanding, are bringing about a "collective good". Notwithstanding that this may offend our epistemic ideals, some interesting lessons can still be derived from this approach, provided we guard against assuming that this sort of cooperation has any epistemic merit.

The Evolution of Cooperation

Cooperation can be studied, not only using game theoretic models such as the infamous Prisoner's Dilemma (Figure 1a)—where agents face a stark clash between what is in their individual best interests and what is best for the group—but also models such as the Stag Hunt (Figure 1b), where mutual cooperation is in everyone's best individual interests, but cooperation makes one vulnerable to the decisions of others, and hence requires substantial trust to be realised (Skyrms 2004). In both cases, there is *something* to be said for withholding cooperation: in the prisoner's dilemma it is straightforwardly beneficial to an individual to withhold cooperation. No matter what others are doing, there is a temptation to be free rider. In the stag hunt, withholding cooperation may lead to reduced benefits for onself, but on the other hand it reduces the risk of being betrayed by others. You want to cooperate when you can count on others to cooperate also, but you want *not* to cooperate otherwise.

	Cooperate	Defect
Cooperate	3,3	0,4
Defect	4,0	1,1

Figure 1a:The Prisoner's Dilemma

Hold fixed what one player chooses, the player always earns a higher payoff by defecting. Defecting is a strictly dominant strategy. The only Nash equilibrium, consequently, is all-defect, notwithstanding that this is a Pareto inferior outcome to universal cooperation.

	Stag	Hare
Stag	4,4	0,1
Hare	1,0	1,1

Figure 1b: The Stag Hunt

Here universal cooperation (Stag, Stag) is again Pareto superior to universal defection, but mutual cooperation is also a Nash equilibrium: neither player has an incentive to unilaterally change strategy from this combination. But the same can be said for the (Hare, Hare) combination, which is also an equilibrium. The challenge to get to the payoff superior equilibrium is to ensure enough trust exists that others will play Stag as well. This challenge is all the more acute in large multiplayer versions of the Stag Hunt.

We know that cooperation is not the default state of the universe: cooperation is fragile, because of its coordinated nature. Evidently, however, cooperation does exist: it exists within species—most obviously in species that collaborate in the rearing of young (Burkart et al 2009, Kramer 2010); it exists across species—such as when a cleaner fish eat parasites from a host fish, which in turn refrains from preying on the cleaner species (Noë and Hammerstein 1995); and it occurs *within* organisms—individuals cells cooperate in elaborate

ways to achieve physiological functions (Aktipis et al 2015). A dramatic example of what can happen when a cell stops "cooperating" in this sense is the development of a cancer.[1]

From modeling and observing these and other examples of cooperation, and its failures, we know that there are a few mechanisms that are able to sustain cooperation. These mechanisms arise in the following circumstances:

> • When the future casts a long shadow. If I stand to be paired with you for a long time, (and you will remember how I behaved in the past) cooperating now is more likely to be in my interests, because cooperating now may bring me benefits of your future cooperation.[2] In these circumstances, cooperation is facilitated by *direct reciprocity*.

> • When cooperators are related. In biology, if we have similar genes, then what is good for your genes is good for mine. So cooperation occurs more readily among family members. By helping your genes, I am helping myself. In non-biological contexts, the same mathematical structure is instantiated if we concerns for each other. If I care about your wellbeing, it is much easier for you to count on me to cooperate with you. This mechanism is known as *kin selection*.

> • When cooperators can earn a reputation. If I don't have repeat encounters with you, but you can affect my future encounters with others by praising or condemning me, then I have more reason to work cooperatively now. This mechanism is known as *indirect reciprocity*, as opposed to direct reciprocity.

All of these mechanisms by which cooperation can evolve require some sort of basic *discrimination*: for direct reciprocity, individuals need to be able to distinguish between those who are likely to be neighbours for a long time versus fly-by-nighters who will be gone in the morning. For kin selection, obviously, individuals need to be able to tell kin from others. And for indirect reciprocity, it is necessary to know whether you have a good or bad

---

[1] One way in which to appreciate the point that cooperation is not the default state of the universe is to focus on the cellular level, and reflect that single celled organisms are much less cooperative, by any sensible means of quantifying, than multi-cellular organisms. Although there are sometimes ingenious instances of cooperation among bacteria such as quorum sensing, these examples are dwarfed by the massive scale of cooperation required between the cells of a bee, for instance, required to make the bee fly and gather nectar, which in turn reflects a complicated cooperative relationship with a plant species, which in turn cooperates with multicellular fungi in order to extract nutrients from the soil, and no doubt many other species too. Although these multicellular phenomena are very salient to us, they represent a small proportion of the total biomass of the earth, and they also represent just the most recent last 1.5 billion years in the history of life. The first two billion years were occupied by single-celled species only. Almost certainly, less cooperative, single celled species will exist for several centuries (at least) after all multi-cellular species have become extinct. In short: less cooperative ways of living are low risk ways of living—they can endure under a wider range of circumstances, because they simply do not depend on the presence of fellow cooperators.

[2] Of course we can't just assume the existence of friendly reciprocators a priori. A more accurate way to put the point is that if reciprocators like this exist, then they can outperform (in terms of fitness) more selfish types who won't be able to sustain a cooperative relationship.

reputation. In all cases, the fundamental requirement is that the cooperative behavior must ultimately be directed towards other agents who have similar genes, and hence are disposed to undertake similar cooperative behaviors.

Wherever cooperation has evolved, then, we can anticipate a corresponding domain of evolutionary pressure. Everyone—both welldoers and illdoers—wants to be *thought* to be neighbours/family/well reputed. And similarly, everyone who is disposed to cooperate wants to avoid being fooled. Indeed, it would be astonishing if cooperation existed, anywhere in the natural world, without a corresponding mechanism to enable agents to be selective about with whom they cooperate.[3]

Lessons for Epistemologists

What lessons might we draw for the study of network epistemology, from these considerations?

First, as mentioned earlier, if we are to exploit the analogy between utility and fitness, it is important that we keep the trust and cooperation quite distinct from epistemic success. The most foolish of conspiracy theories and hokum may be sustained in networks that are thick with trust and cooperation. Other perspectives may be able to derive fruitful results using a more normatively loaded conception of trust and/or cooperation, but we need to guard against transferring connotations from those approaches to the present context.

Second, having drawn attention to the importance of mechanisms that discriminate between those one does and does not wish to cooperate with, I suggest that this is a dimension of online epistemic behavior that warrants more attention. Retweeting a rumour may strike us, as epistemologists, for its pernicious epistemic character. But for the tweeting agent it may be serving a much more important *signaling* function than its epistemic effect. Consider many of the topics that are discussed among conspiracy theorists: 9/11, Jews, Islam and race; police, weapons and the military; world leaders and government agencies.[4] It is doubtful that many practical decisions are being made which turn on the sort of testimony offered on these topics. On the other hand, if you turn to a subreddit on DIY advice, there will be testimony about timber and glue products, woodworking techniques, where to find other sources of information, and so forth. There are substantial pragmatic stakes involved in accepting/rejecting testimony about these latter topics. If the testimony is good and is accepted, people will enjoy more success in their DIY projects. Whereas in conspiracy

---

[3] Of course, sometimes mistakes are made. Some cuckoo species, for instance, are brood parasites—they lay their eggs in other birds' nests, exploiting the nesting and rearing behavior of another species, to have their young reared for them. While there is a very strong evolutionary pressure on host species to raise their own young diligently, there is also pressure to avoid being tricked. Presumably the adaptations required to make more precise discriminations among candidate eggs are too far away in genetic space to enable many species to develop suitable defences. (It is entirely possible of course, that some species already have evolved to defend against this parasitism, and that cuckoos have coevolved not to lay their eggs in nests of those species.)

[4] This list is inspired by the topic modelling in Klein et al 2018.

theorising, people could accept or reject all the proffered testimony, and still enjoy roughly the same levels of practical success in their daily lives.

This disconnection from pragmatic consequences means that adopting or espousing a false belief in these domains is not especially costly. The investment of effort in such behavior therefore may not be substantially motivated by a concern to establish the truth. I conjecture that such testimony may, for some individuals, be serving a signaling function. By adopting beliefs that are controversial but which are politicized in some way, an agent may be able to convince others that they are similar to them, and hence contribute to the existence of a trusting relationship that has other, more pragmatic dimensions. When we hear that our neighbour believes that Obama was born in Kenya, we do not marvel at how the neighbour manages to function in the world, but we do draw some inferences about whether or not they were likely to vote for Trump in 2016.

This conjecture is motivated by findings in social science and psychology.

> • Morally charged information is more likely to be shared using social media, and its diffusion is heavily influenced by political associations (Brady et al 2017, 2018). If epistemic concerns were of the greatest import, why would this information be disproportionately likely to spread, and why would its spread be affected by its political orientation. But if the testimony is serving a function of signaling one's affiliation to a cultural identity group, this is a readily explicable pattern.

> • In areas where there is an established correlation between cultural identity and repudiation of scientific discourse—such as safety of vaccines, climate science, and the theory of evolution by natural selection—there is evidence that more scientifically literate individuals are *more* likely to deny the scientific consensus (Kahan 2015, Levy 2019). This is hard to understand as reflective of any epistemic goal, but can be explained if the denial of scientific assertions in these domains is an important marker of cultural identity: more knowledgeable individuals have to "work harder" to assert their identity here by their denial of scientific authority.

This evidence of course is not enough to exclude other possibilities, nor do I deny that there is almost certainly substantial variation across individuals in motivation for engaging in this behavior. One stereotype of the most ardent conspiracy theorist, for instance, is somebody who has a very great concern to establish the truth, mixed with pathological attraction to contrarian or eccentric theories, and an ignorance or indifference to the adverse social consequences of adopting such ideas—a signaling function is rather less plausible in that instance. But for at least some individuals, signaling is a hypothesis at least worth further investigation.

The signaling hypothesis leads to several predictions, which may be tested empirically.

> 1. If the signals are relatively cheap, they will either be relatively meaningless, or they will be used in communities that have tightly aligned interests.
>
> 2. If the signals are relatively costly, they may be used for effective communication even between individuals with opposed interests (Gambetta 2009).

Similar ideas relating to the cost of a signaling device have been explored in the context of religious communities and the relative burden of complying with the religious doctrine. Lawrence Iannaccone (1992) argues that by requiring costly sacrifices such as forgoing alcohol, sex, popular entertainment, the wearing of stigmatizing clothing, etc, a group will be able to screen out less pious free-riders—types who enjoy some of the benefits of organised religious activity but are less willing to contribute the effort required to sustain that activity. This model purports to explain why strict religious orders may be both small (not everyone is willing to pay the "costs" of participation) but also robust and enduring (those who do join are highly committed, good contributors). In the sorts of cases that I am describing the cost of adopting a relatively politicized conspiracy theory is much less costly. While it may stigmatize one with regard to the other side of the political debate, in a largely binary political environment, it still leaves a large group of allies. Iannoccone's model predicts therefore that the average level of contribution to groups sustained by such screening devices is likely to be much lower than in the devout religious groups he describes. This is a promising area for further investigation.

Conclusion

In this short piece I have used the framework of evolutionary biology to, first, emphasise the relatively special nature of cooperation, as a natural phenomenon, and argued that online networks are an instance of cooperation, so understood. Further, drawing on the fact that cooperation cannot be stable unless it is underscored by some sort of discrimination between cooperators and cheats, I have conjectured that much online behavior that is overtly testimonial may in fact also be serving a signaling function. The cooperation may be overshadowed by the efforts to distinguish different varieties of cooperative partners from each other.

Proposing a signaling hypothesis for a given variety of online testimony is not to say that such behavior has no epistemic character. It can be both a signal and an epistemic behavior. But because the two dimensions are intimately associated, it is difficult to theorize optimal epistemic dispositions in isolation from questions about how our epistemic networks should be constituted.[5] And because of this, it bolsters the case for using tools like agent-based models to study social epistemology, rather than more traditional thought-experiment based methods. Thought experiments are rarely able to accurately interrogate what the result will

---

[5] This is arguably another manifestation of what Kahan 2015 calls the "measurement problem" in the science of science communication. Attempts to measure individual beliefs may be contaminated by the degree to which they instead measure cultural affiliations, and vice versa.

be of two dynamical processes interacting simultaneously, and that is precisely what appears to be occurring in the online networks that concern us.

References

Aktipis, C. Athena, Amy M. Boddy, Gunther Jansen, Urszula Hibner, Michael Hochberg, Carlo Maley, Maley, and Gerald S. Wilkinson. 2015. "Cancer Across the Tree of Life: Cooperation and Cheating in Multicellularity." *Philosophical Transactions of the Royal Society B: Biological Sciences* 370 (1673): 20140219. https://doi.org/10.1098/rstb.2014.0219.

Alfano, Mark and Colin Klein. 2019. "Trust in a Social and Digital World." *Social Epistemology Review and Reply Collective* 8 (10): 1-8. https://wp.me/p1Bfg0-4tk.

Baier, Annette. 1986. "Trust and Antitrust. *Ethics* 96 (2): 231–260.

Brady, William J., Jay Bavel, John Jost, and Julien Wills. 2018. "An Ideological Asymmetry in the Diffusion of Moralized Content Among Political Elites." https://doi.org/10.31234/osf.io/43n5e

Brady, William J., Julian A. Wills, John T. Jost, Joshua A. Tucker, and Jay J. Van Bavel. 2017. "Emotion Shapes the Diffusion of Moralized Content in Social Networks." *Proceedings of the National Academy of Sciences* 114 (28): 7313–7318. https://doi.org/10.1073/pnas.1618923114.

Burkart, Judith M., Sarah B. Hrdy, and Carel P. van Schaik. 2009. "Cooperative Breeding and Human Cognitive Evolution." *Evolutionary Anthropology: Issues, News, and Reviews* 18 (5): 175–186. https://doi.org/10.1002/evan.20222.

Gambetta, Diego. 2011. "Signaling." In *The Oxford Handbook of Analytical Sociology* edited by Peter Bearman and Peter Hedström, 168–194. doi: 10.1093/oxfordhb/9780199215362.013.8

Kramer, Karen L. 2010. "Cooperative Breeding and its Significance to the Demographic Success of Humans." *Annual Review of Anthropology* 39 (1): 417–436. https://doi.org/10.1146/annurev.anthro.012809.105054

Iannaccone, Laurence R. 1992. "Sacrifice and Stigma: Reducing Free-Riding in Cults, Communes, and Other Collectives." *Journal of Political Economy* 100 (2): 271–291. https://doi.org/10.1086/261818.

Kahan, Dan M. 2015. "Climate-Science Communication and the Measurement Problem." *Advances in Political Psychology* 36 (Suppl. 1) 1–43.

Klein, Colin, Peter Clutton, and Vince Polito. 2018. "Topic Modeling Reveals Distinct Interests Within an Online Conspiracy Forum." *Frontiers in Psychology* 9: 189. https://doi.org/10.3389/fpsyg.2018.00189.

Levy, Neil. 2019. "Due Deference to Denialism: Explaining Ordinary People's Rejection of Established Scientific Findings." *Synthese* 196 (1): 313–327. https://doi.org/10.1007/s11229-017-1477-x.

Maynard Smith, J. 1974. "The Theory of Games and the Evolution of Animal Conflicts." *Journal of Theoretical Biology* 47 (1): 209–221. doi: 10.1016/0022-5193(74)90110-6.

Noë, Ronald and Peter Hammerstein. 1995. "Biological Markets. *Trends in Ecology & Evolution* 10 (8): 336–339. https://doi.org/10.1016/S0169-5347(00)89123-5.

Skyrms, Brian. 2004. *The Stag Hunt and the Evolution of Social Structure.* Cambridge University Press, Cambridge.