



# The coordination dilemma for epidemiological modelers

Ignacio Ojea Quintana<sup>1</sup> · Sarita Rosenstock<sup>1</sup> · Colin Klein<sup>1</sup>

Received: 17 August 2020 / Accepted: 10 November 2021 / Published online: 26 November 2021  
© The Author(s), under exclusive licence to Springer Nature B.V. 2021

## Abstract

Epidemiological models directly shape policy responses to public health crises. We argue that they also play a less obvious but important role in solving certain coordination problems and social dilemmas that arise during pandemics. This role is both ethically and epistemically valuable. However, it also gives rise to an underappreciated dilemma, as the features that make models good at solving coordination problems are often at odds with the features that make for a good scientific model. We examine and develop this dilemma in the context of the COVID-19 pandemic, and suggest extensions to other domains.

**Keywords** COVID-19 · Pandemic · Game theory · Coordination · modeling

## Introduction

### Coordination and public signals

Collective action often requires determining both the best thing to do and how to get everyone to do the same thing. The two demands can come apart. A group of friends with diverse preferences wants to go to dinner. They may all prefer eating anywhere to not eating at all, yet the hunt for the best restaurant (conditional on their preferences) may take so much time that they end the evening separate and hungry. Conversely, as Schelling (1960) famously noted, coordination on a satisfactory solution may be achieved by otherwise arbitrary means, so long as they are jointly salient. One friend shouts ‘Pizza!’ and walks off: that is good enough to force coordination and saves hours of wrangling. Indeed, the power of salient coordination points can be so strong that they can lead to otherwise undesirable stopping points: a well-timed radio ad might lead the group to the dodgy buffet, even though they all barely prefer it to starvation.

---

✉ Ignacio Ojea Quintana  
ignacio.ojea@anu.edu.au

<sup>1</sup> School of Philosophy, The Australian National University, Canberra, Australia

Considerations of coordination also apply to public health. Epidemiological modeling, and its communication to the public, plays an important role in finding and justifying the right course of action during public health crises. Such models are also involved in what Goodin (1980) calls “the politics of the obvious.” By presenting a course of action as sensible, each person can also have confidence that everyone else has seen the same signal of what the generally determined sensible course should be. Even if one thinks that the sensible is wrong, the fact that it is obvious to all carries its own rational force.

This common signal can solve coordination problems. Suppose that a certain vaccination program is only worth it if everyone (or enough people) do it, and that I would only prefer to be vaccinated if everyone else does. A public signal that vaccination is good might move me to get vaccinated, precisely because I think that it provides a salient coordination point that will get everyone else moving. The converse can happen as well. A botched ad campaign might lead me to believe that people will be unconvinced, and so fail to get vaccinated, which it turn lowers my willingness to do—even if this is not what I would prefer all things considered.

### The coordination dilemma

These dynamics can lead to an underappreciated dilemma faced by epidemiological modelers, one that derives from the role of models in communicating policy choices. As the dilemma became particularly acute at several points during the COVID-19 pandemic, we use it as a case study.

A *coordination dilemma* arises because the public communication of epidemiological models also solves certain kinds of coordination problems. Many pandemic responses require everyone to do the same thing; even the sub-par coordinated outcomes are often better than the best uncoordinated ones. Yet insofar as particular models are made public at all, that coordination function becomes obligatory: by providing salient coordination points, modelers should expect that what they do will promote or undermine coordination. Given that, modelers face a dilemma: scientific fidelity may require caution and care, but coordination requires a strong signal. So good modeling may make for bad outcomes, and vice versa. Note that ‘public’ here is a matter of availability rather than intention: as we discuss in the final section, the dilemma becomes more acute the more studies are (e.g.) shared on social media.

To illustrate the dilemma, consider a real life example. In March 2020, the Imperial College London made public a report in which they argued that

...optimal mitigation policies (combining home isolation of suspect cases, home quarantine of those living in the same household as suspect cases, and social distancing of the elderly and others at most risk of severe disease) might reduce peak healthcare demand by 2/3 and deaths by half.

Their arguments were based on several modeling and simulation assumptions, yet it had a significant public impact both in the general public and the policy makers. Ten days after this first report, a group of academics at Oxford published an article Lourenço et al. (2020) using similar modeling techniques. Among many nuanced

points, the report claimed that there were already “significant levels of herd immunity in both countries.”

On the face of it, the two reports were in conflict. One suggested the need for strong public health responses, the other suggested that these would be unnecessary. This is an ordinary, and often unremarkable feature of scientific modeling in uncertain domains: good models disagree, and it can take time to sort out the right answer. Yet this made the public signal about how people should behave, and could expect others to behave, much more obscure. People reading the two reports, even if they were convinced that one was correct, could not have confidence that collective action would ensue.

A coordination dilemma arises when the demand to do good science creates conflicts with the role that the scientific outputs play for coordinating our actions.

## The plan

The goal of this paper is to articulate this coordination dilemma, formalize it to show the connection to classic coordination problems studied in other domains, and suggest specific conditions under which it arises. The astute reader will note that that list does not contain anything about *solutions* to the dilemma once it arises. The benefit of formalization is to show that this is really a dilemma. The best we can hope for is to avoid the dilemma in the first place. We therefore conclude with reflections on how one might avoid the coordination dilemma when it might otherwise arise.

## Distinguishing the dilemma

We begin by distinguishing the coordination dilemma from what we call the standard dilemma of policy-relevant science.

Philosophers concerned with the role of science in policy decisions often do so in terms of *inductive risk*. As Rudner (1953) puts it, “How sure we need to be before we accept a hypothesis will depend on how serious a mistake would be”. No matter how good your evidence is, any scientific conclusion will leave *some* room for error. The risks associated with potential errors are compounded by the weight given to scientific judgments in public policy (Douglas 2009).

The lens of inductive risk frames the normative responsibilities of scientists in terms of the veracity of the contents of their claims. This involves balancing certainty of scientific conclusion with the urgency of action, in light of expectations of what would happen if the conclusion is or is not correct. It also may involve effectively communicating the strength of evidence and degree of uncertainty. In contrast, the coordination dilemma is somewhat orthogonal to the question of whether we come to the ‘right’ conclusions about how the world is. It is neither necessary nor sufficient to resolve the coordination dilemma for us to be ‘right’. As in the case of friends deliberating on where to go for dinner, our chances of correctly identifying

the optimal path forward are slim, and our ability to benefit from *any* choice is tied to our ability to promptly achieve some sort of consensus.

Work on *science communication* looks beyond the truth conditions of scientific claims and considers their material presentations as rhetorical acts. This literature considers how non-factive features of science communications such as framing, curation, and timing, impact how the scientific results are attended to, comprehended, and used to motivate action by individuals and policy-makers (Medvecky and Leach 2019).

While more suited to our purposes, this literature still tends to frame the normative role of science communicators as concerning the epistemic and behavioral effects of science communications on individual consumers of those artifacts. Understanding the coordination dilemma requires more than just considering how an individual interacts with a scientific output. It also requires considering how collections of individuals conceptualize and interact with one another with respect to those outputs. We will be concerned with cases where an individual on their own would behave differently than an individual who knows that they are part of a group.

The line between science and science communication is not sharp. The increasing tendency of scientists and modelers to publish works in progress on sites like arXiv, and communicate directly with the public on social media is exacerbated by the public's need and demand for COVID-19 research and information, and renders more and more of scientific activity as acts of public communication. We thus do not attempt to draw a distinction, but only refer to the aggregate outputs of scientific work.

For the coordination dilemma to occur, these outputs must be understood (implicitly or explicitly) as supporting an argument for bringing about a particular state of affairs that can only be achieved by the cooperation of individuals. To express the dilemma, we draw a distinction between first- and second-order contents of the scientific outputs.

The *first-order* content is the (implicit or explicit) argument itself. Considering it requires a straightforward evaluation of the validity of the argument. For a layperson, may largely be a question of rhetoric, or how convincing the argument is to them. It can be made arbitrarily more sophisticated, however, by roping in tools from epistemology and philosophy of science to assess the strength of evidence, as well as considerations of epistemic risk.

The first-order content alone suffices to understand what we call the *standard dilemma* of policy-relevant science. This dilemma occurs when the norms of good scientific practice conflict with the rhetorical efficacy of the first-order message. Good scientific practice (including modeling work) requires a cautious appraisal of alternatives, the quantification of uncertainty, and the presentation of multiple points of view. Yet uncertainty, ambiguity, and multiplicity undermines the rhetorical punch of the first order message. Conversely, strong, overly simple, crude modeling can make be more rhetorically effective, even though that might undermine the scientific justification motivating the first-order signal itself.

Of course, even outside of the public eye, scientists already have to make various trade-offs among purely scientific virtues. These include balancing the explanatory benefits of simple, general models with the richer, possibly more

accurate picture they might get from adding more complexity to their models. Considerations of rhetorical strength bring about a meaningfully different trade-off, which need not align with that of the scientist. Perception that models are overly simplified may lead people to expect others to discredit them as unrealistic. And conversely, models that are complex enough to muddy the causal narrative may also weaken its persuasiveness.

The standard dilemma is related to but distinct from the focus of our paper, the coordination dilemma, which involves considering what we call second-order content as well. The *second-order* content concerns whether and how the first-order content can serve as a salient coordination point. One might find modeling results about the efficacy of social distancing compelling, but nonetheless doubt that their peers will find it similarly compelling, thus rendering the modeling results inapplicable to their situation.

The *coordination dilemma* arises when the norms that govern the scientific practices generating first-order content conflict with the coordination role played by the second-order content. That is, even once we have solved the standard dilemma and enacted best practices for generating first order content—reliable scientific advice presented as compellingly as possible without sacrificing rigor—the result may still fall short of what is needed for scientific messaging to serve as a touchstone for coordination.

This is because attempting to address the standard dilemma without considering second order content may leave out key features of the social context in which the scientific message is translated into action. A common refrain regarding the standard dilemma, for example, is to promote better scientific literacy in the public. If an individual has internalized the norms of good scientific arguments, then what they find rhetorically compelling will be more closely aligned with the outputs of good scientific practice. But unless “scientific literacy” is achieved universally simultaneously, the coordination-dilemma still looms. Scientific literacy may even make an individual *more* sensitive to second-order concerns. The more critically one examines the assumptions underlying a scientific inference, the more weak spots they will inevitably find where they may worry that public doubts will fester.

Conflicting values can also undermine the second-order message. If I think that mask mandates convey a socialist set of values that are incompatible with my rugged individualism, that is not yet a reason to think that the mandates are wrong, or even that I ought not to comply. But if I think that enough others share my values, and find a similar conflict with the assumptions of the first-order models, then I might reasonably think that enough people won’t comply for compliance to be worth the costs.

Addressing the coordination dilemma likely requires intervening on cultural and interpersonal expectations, not merely individual values and epistemic standards. In this way, it is analogous to the electability dilemma for politicians—messaging that makes a politician appealing to an individual may conflict with messaging that convinces that individual that their appeal is widespread enough to merit support.

## Models, public signals, and games

The coordination dilemma arises when the demand to do good science creates conflicts with the second-order role that the scientific outputs play for coordinating our actions. With that in mind, we now turn to game theory to provide a more clear representation of the problem. On the one hand, this requires developing the type of scenarios that require collective coordination for their best response. We will consider coordination games and social dilemmas (with punishment) for their representation, and we will argue that pandemic response can be reasonably modeled this way. On the other hand, we need an understanding of the role that public signals have in those scenarios. For this we will make use of Aumann (1987) and the notions of *correlated equilibria* and *correlating devices*. The upshot of the section is putting both pieces together and showing that noisy public signals cannot secure the best outcomes. In particular, even if the first order recommendation is correct, if the second-order message is vague, agents might not comply with the recommendation.

A normal form *game* consists of

- (i) a set of players  $i = 1, 2, \dots, n$ ;
- (ii) strategy/action sets  $S_1, S_2, \dots, S_n$  for each player ( $S = S_1 \times S_2 \times \dots \times S_n$ ); and
- (iii) a payoff/utility function for each player  $i$ ;  $u_i : S \rightarrow \mathbb{R}$ .

A *mixed strategy* for player  $i$ ,  $\sigma_i \in \Delta(S_i)$ , is a probability distribution on  $S_i$ . Mixed strategies can be thought as each player having an independent randomizing device and defining their actions as conditional on the outcome of that device. This is not the canonical interpretation, but it helps in anticipating the notion of correlating device that we will use later to model public signals.

Solutions to games are strategy profiles with an equilibrium-like property. We can think of strategy profiles  $(\sigma_1, \dots, \sigma_n)$  as having two components for each player  $i$ : (i) a (first-order) suggestion on what they should do ( $\sigma_i$ ), and (ii) a (second-order) expectation of what others will be doing ( $\sigma_{-i}$ ).

In general, a strategy profile will have an equilibrium-like property if (i) and (ii) are compatible with maximizing expected utility, namely if conditional on (ii) players would rather not deviate from (i). The best example of a way to have this property is the Nash Equilibrium (NE). A (possibly mixed) strategy profile  $(\sigma_1, \dots, \sigma_n)$  is a NE of  $G$  if for every  $i$ , and every  $s_i \in S_i$ :  $u_i(\sigma_i, \sigma_{-i}) \geq u_i(s_i, \sigma_{-i})$ . In the mixed strategy case, each player is *certain* about how others will act as a function of the outcome of their independent randomizing devices; and conditional on *that*, they would rather not deviate from their own strategy.

## Public signals and correlated equilibria

By contending that epidemiological models serve the role of public signals we mean that in their prescriptive and informative role they (a) suggest a course of action for individuals, and (b) they give each individual an expectation on whether others will

follow their course of action—i.e., how salient the solution is. This is a formalization of the distinction we developed before between a first-order and a second-order message.

In what follows, we conceive of the relevant “signals” to be any communicative artifact based on modeling work that is situated in an implicit or explicit argument for a particular course of action. We argue that such public signals can be conceived as correlating devices that can be useful in coordinating collective action. This work builds on the notion of *correlated equilibria* from Aumann (1987).

Before introducing the concept formally, let us build on our original example a group of friends coordinating on a restaurant. Suppose there are six friends with diverse preferences, so that each friend has a distinct preferred restaurant and gets the best reward if they all go there, a second best reward if they go together to another restaurant, a third where they go to their favorite restaurant but not together, and zero payoff if not only they fail to coordinate but also ends up in a non-preferred spot. One reasonable solution for them is to use a (shared) fair die as a correlating mechanism, and publicly commit to go to the restaurant associated the outcome of the die. After observing the outcome of the die, and assuming others will fulfill their commitments, they each expect with certainty all others to go to its associated restaurant, and on that expectation they would rather comply with their own commitment. This is a correlated equilibrium.

But suppose now that the correlating device is noisy. Although the outcome of the die is visible to all, imagine that the strategy they committed to was *not* publicly announced. So each agent made a commitment (the same as before) and can observe the outcome, but they are uncertain about the commitments of others. Given this uncertainty, individuals cannot reasonably expect successful coordination, and might prefer to go to their favorite place alone regardless of their personal promise. In that case this strategy profile is not a correlated equilibrium.<sup>1</sup>

Scenarios like our example of the group of friends can be easily formalized with these tools.<sup>2</sup> Correlated equilibria are conceptually rich. In his seminal paper Aumann (1987) shows that under the assumption of common priors, if all players are Bayes rational (i.e. they maximize expected utility conditionalizing on the information), the resulting strategy tuple will be a correlated equilibrium. Furthermore, correlated equilibria are the more general equilibrium concept in game theory, with other notions like Nash, evolutionary stable, sub-game perfect, etc. being refinements. More relevant to our purposes here, correlated equilibria have been proposed

<sup>1</sup> Formally: A *correlating mechanism* is a triple  $(\Omega, \{H_i\}, p)$  where  $\Omega$  is a set of states,  $p$  a probability distribution on  $\Omega$ , and for each player  $i$ ,  $\{H_i\}$  is an information partition of  $\Omega$ . A *correlated strategy* for a player  $i$  is a function  $f_i : \Omega \rightarrow S_i$ , that is measurable with respect to  $\{H_i\}$  (i.e. it assigns the same action to elements of the same partition component). A strategy profile  $(f_1, \dots, f_n)$  is a *correlated equilibrium* relative to mechanism  $(\Omega, \{H_i\}, p)$  if for every  $i$  and correlated strategy  $g_i$ ,

$$\sum_{\omega \in \Omega} u_i(f_i(\omega), f_{-i}(\omega))p(\omega) \geq \sum_{\omega \in \Omega} u_i(g_i(\omega), f_{-i}(\omega))p(\omega).$$

<sup>2</sup> In the equilibrium case,  $\Omega = \{1, 2, 3, 4, 5, 6\}$ ,  $p$  is uniform,  $H_i$  is the finest partition for all  $i$ , and  $f_i : n \mapsto i$  denotes that person  $i$  goes to restaurant  $n$ . The non-equilibrium case we can an model it using a *direct mechanism* in which  $\Omega = S = S_1 \times S_2 \times \dots \times S_6$ ,  $p$  is uniform, and  $S_i = \{1, 2, 3, 4, 5, 6\}$  and  $H_i = \{s' \in S : s'_i = s_i\}$  for all  $i$ .

to model conventions (Vanderschraaf 1995), and more broadly the emergence and evolution of social norms (Morsky and Akçay 2019) and other social devices (Skyrms 2014). Correlating mechanisms provide a clean formalization of Schelling (1960)'s notion of *focal point* and Lewis (1969)'s idea of *salience*. We use these notions to model *public signaling*.

Public signaling is a well understood mechanism to secure success in coordination (and other) games (Skyrms 2010). For our purposes here, a public signal corresponds to a profile of correlated strategies relative to a mechanism. In other words, it both recommends an action to players (i.e. the first-order message) and gives them an expectation of what the other players will end up doing (the second-order message) relative to a correlating mechanism. For example, if by surveying academic modeling of the COVID-19 contagion, social distancing is consistently recommended for everyone, then our player takes it as a suggestion but also expects most people to comply.

There are many ways in which public signals help coordination. Most obviously, public signals make some actions more *salient* and therefore more appealing to players. Salience can often be sufficient to break ties when there are multiple good equilibria (Schelling 1960). Even if coordination is enforced by punishment, effective communication of the *threat* of punishment is an important part of the scenario.

In this section we begin by modeling pandemic response as a coordination game. We proceed to demonstrate how the situation becomes more challenging when it instead takes the form of a social dilemma. For both cases, we show how noisy public signaling can be detrimental to obtaining the best response.

## Public signaling in a coordination problem

We argued that some circumstances lead to tensions between the kind of broader scientific demands that constrain (epidemic) modeling and the role that those models play in public signaling. In the next sections we will present two natural strategic representations of those circumstances, coordination games and social dilemmas, and show how a (possibly scientifically accurate) signal might fail to become a successful coordination point if its second order message is not clear.

### Coordination games

In *coordination games* the central problem is to settle on one of the many distinct Nash equilibria in the game. While there is a wide class of interesting coordination games, we focus here on the *Stag Hunt* (SH), as it is relatively simple to present and analyze.

In a stag hunt, two players each have the option of hunting either a stag or a hare. The stag offers more food, but is a riskier choice since it will be much more difficult to hunt if the other player does not choose to pursue it as well. This situation is presented formally in Table 1.

SH can be generalized by, e.g. increasing the number of players or by grading the levels of cooperation between Stag and Hare. SH has three Nash Equilibria:



**Table 1** Symmetric Coordination Game or Stag Hunt:  $A > B \geq D > C$

	Stag	Hare		Stag	Hare
Stag	A, A	C, B	Stag	4, 4	1, 3
Hare	B, C	D, D	Hare	3, 1	2, 2

The entries in each row and column correspond to the pay-offs for the “Row” player and “Column” player respectively

two pure ones and a perfectly mixed strategy for each player with payoff  $2\frac{1}{2}$  for each. Important for our purposes, (Hare, Hare) *risk dominates* (Stag, Stag), a property explored by Harsanyi and Selten (1988). Risk dominant profiles are attractors in some dynamics like the replicator dynamics; but our target interpretation is one concerning uncertainty. The intuition behind risk dominance is that it is safer for each player to play ‘Hare’, if they are uncertain about what the other player will do.<sup>3</sup> Hence in the face of uncertainty about what others will do, the problem of coordination games is that players might end up in a sub-optimal equilibrium. As the reader might expect, these are the kinds of situations that emerge if public signals are not sufficiently strong.

### Pandemic response as a coordination problem

We contend that the practice of “social distancing” as a response to the COVID-19 pandemic admits a strategic representation either as a coordination game or as a social dilemma. In both cases the use of idealized models can serve strategic purposes that go beyond their scientific and explanatory reach.

“Social distancing” refers to a collection of strategies used by individuals to limit exposure during a pandemic. We use it in a broad sense to include specific actions like physical distancing or wearing surgical masks, as well as strategies that involve limiting the connectivity of social graphs by avoiding symptomatic patients or avoiding most contacts altogether. Behavioral influences on pandemics have long been known, but the past twenty years have seen a concerted effort to model their effects (Funk et al. 2010). Importantly, in much of this literature it is assumed that social distancing can be driven by entirely bottom-up factors: rumors about the spread of a disease in a local community can be sufficient to get people to begin social distancing, for example (Funk et al. 2009).

That individuals should and do avoid contacts with the obviously symptomatic is unsurprising. However, large-scale epidemics with asymptomatic transmission and widespread community spread, more proactive and more serious social distancing is necessary to slow the spread of the outbreak. We assume that agents have a choice between distancing or continuing to engage in social activities. We also assume that

<sup>3</sup> Suppose Row expects Column to play Stag with probability  $p = 0.4$ , then their expected payoff for hunting the stag is  $E(\text{Stag}) = 0.4 \cdot 4 + 0.6 \cdot 1 = 2.2$ , while their expected payoff for hunting the hare is  $E(\text{Hare}) = 2.4$ .

social distancing brings with it some cost—the loss of ordinary social contacts and activities is a burden.

There are two focal strategy profiles in social distancing: either a sufficient number of people do it or nobody does it. We assume, following Reluga (2010)'s game-theoretic modeling of social distancing, that social distancing bears some costs “in terms of liberty, social capital, time, convenience, and money,” Any middle-ground would fail to effectively reduce the spread of contagion significantly, which would lead individuals to abandon the costs of social distancing.<sup>4</sup> These focal profiles can be seen as either both equilibria in a coordination problem, or as a social dilemma, depending on how we conceive of the payoffs in the case where everyone distances.

### Public signaling in a coordination problem

Consider first the case in which widespread social distancing is an effective means to control pandemics, and that every individual is better off in this scenario. This means that an equilibrium in which the great majority of individuals observes distancing has individually and collectively higher payoff than one in which (almost) no one does. Everyone would be better off if we managed to hold the social distancing policy for a few months and avoid the death of tens of thousands of vulnerable individuals. Considered as such, the question of social distancing resembles games like stag hunt, with a payoff dominant equilibrium if nearly everyone distances.

It would be naïve to neglect the alternative profile. The equilibrium in which almost no one practices social distancing seems to be risk dominant (Harsanyi and Selten 1988), in the same way as ‘Hare’ was risk dominant before. If an individual expects others to continue to socialize, then they might have an incentive to socialize too. It does not make much sense to continue to distance while others do *not* do so; staring at a neighbor’s party from a solitary balcony is not fun. If everything is fine, then visiting *your* friend poses no threat to anyone. On a more strategic level, competing business and competing nations might feel pressure to relax distancing if they observe their neighbors not doing so. Consider the replicator dynamics, where the number of initial hare-hunters necessary to generate a cascade of defection (i.e. other hare-hunters) depends on the distance between  $A$  and  $C$ . With the current values ( $A - C = 4 - 3 = 1$ ) it is enough to have  $> \frac{1}{2}$  of the population, with value  $A - C = 0.5$  it is sufficient to have  $> \frac{1}{3}$ . In other words, it can sometimes take a minority of hare-hunters to initiate a cascade of hare-hunters. The hare-hunting equilibrium is an attractor.

Let us return to *Stag Hunt* and briefly consider how signaling operates as a correlating devices there.

To study correlating equilibria it is sufficient to consider *direct mechanisms* in which  $\Omega = S = S_1 \times S_2$ ,  $S_i = \{\text{Distance, Not-Distance}\}$  and  $H_i = \{s' \in S : s'_i = s_i\}$

<sup>4</sup> Reluga (2010)'s model shows that the threshold for ‘sufficiently many’ depends on both the basic reproductive number  $R_0$  and the overall efficiency  $m$  of social distancing; see esp. Figs. 4 and 5 of his (2010). The perceived best strategy also depends on expectations about time to future vaccination availability—another realm in which the modeler’s dilemma may become relevant.

**Table 2** Left: stag hunt-symmetric coordination game form. Right: correlating mechanism distributions

	Distance	Not-Distance		Distance	Not-Distance
Distance	4, 4	1, 3	Distance	$p_1$	$p_2$
Not-Distance	3, 1	2, 2	Not-Distance	$p_3$	$p_4$

for both Row and Column, and  $p$  is given by  $p_1$  to  $p_4$  ( $\sum p_i = 1$ ). For simplicity and to preserve symmetry assume  $p_2 = p_3 = p^*$ , which in this setup just means that everyone shares the same prior expectations that others will *not* distance if they distance (or distance if they do not distance).

Consider a nuanced epidemiological model which, while stating the limitations of the study that motivates it, provides a clear first-order directive that individuals should exercise social distance. Suppose also that you, the reader, find the suggestion sensitive and limitations stated an expression of honesty that do not debunk the central claim that social distancing is the best response to the pandemic. Yet you are uncertain about how that suggestion, with the study limitations, will be evaluated by other members of society. The second-order message is noisy. In our context this means that  $p^*$  is relatively large. If  $p^*$  is large enough, then the signal will not be a correlated equilibrium, and the more pressing the coordination problem, the *stronger* (less noisy) the signal needs to be in order for it to be strategically consistent.<sup>5</sup>

These precise values do depend on the particular parameterization of the problem we proposed, but the main take-away that public signal uncertainty might lead to deviations from optimal coordination still hold. The upshot is that in behaving like public signals, epidemiological models can not expect the audience to (rationally) follow their suggestion if the second-order message is sufficiently noisy. But as explained before, reducing the noise of the second-order message might be in conflict with scientific modeling practices.

### Public signaling in a social dilemma

#### Social dilemmas

Another type of scenario in which the coordination dilemma might emerge are *social dilemmas*. We will now briefly introduce them and discuss how pandemic response can be conceived as such.

Unlike *coordination games*, in which the issue is how to settle on one of several Nash equilibria, in *social dilemmas* all Nash equilibria are somewhat dissatisfying. Social dilemmas are characterized by the existence of some non-equilibrium strategy profile for which each individuals' payoff (or more weakly, aggregate welfare)

<sup>5</sup> For example, suppose that in the parameterization in Table 2, we changed  $B = 3$  to  $B = 3.5$  so that players have more of an incentive not to distance if others are. Then if  $p^* > \frac{2}{5}$  the public signal would not be a correlated equilibrium.

**Table 3** Prisoner’s Dilemma: In general its only required that  $T > R > P > S$ , and sometimes also that  $2R > T + S$

	Cooperate	Defect		Cooperate	Defect
Cooperate	R, R	S, T	Cooperate	3, 3	1.5, 3.5
Defect	T, S	P, P	Defect	3.5, 1.5	2, 2

**Table 4** Two Person Three Actions Public Goods Game: If we were to eliminate one of the actions for both players, we get a Prisoner’s Dilemma

	Pool 2	Pool 1	Pool 0
Pool 2	3, 3	2.25, 3.25	1.5, 3.5
Pool 1	3.25, 2.25	2.5, 2.5	1.75, 2.75
Pool 0	3.5, 1.5	2.75, 1.75	2, 2

is better than in any Nash equilibrium profile. The obvious and simplistic example is Prisoners Dilemma (PD), presented formally in Table 3.

In PD, each of two prisoner’s has the option to either ‘Cooperate’ with the other and stay silent, or ‘Defect’ and incriminate their partner in exchange for a lighter sentence. Each is best off when they defect while their partner cooperates, but these strategy profiles are not NEs, as cooperators will be better offer switching to defection. Mutual cooperation is also not NEs, as each partner stands to benefit from unilaterally defecting.

Thus the only NE is mutual defection, despite being payoff-dominated by mutual cooperation. As with SH, PD can be generalized in many ways, still preserving the social dilemma properties mentioned before. For example, an  $n$ -person PD is one in which  $n$  individuals can choose whether to cooperate or defect. Their payoffs follow the PD distribution, but controlling by proportions.<sup>6</sup> Given that  $T > R > P > S$  it is easy to show that general defection is the only Nash equilibrium. Another generalization of PD is the Public Goods (PG) game, in which agents have degrees of cooperation and defections. In PG, each agent has a budget and then can choose to invest as much as they want from that budget into the public pool, then the pool is increased by a certain factor ( $> 1$  and less than the number of players), and equally distributed among players. A simplistic two person, \$2 budget, and 1.5 factor is shown in Table 4. This version can be generalized by admitting more graded cooperation and by incorporating more agents as we did before with PD.

The social dilemma concern is still preserved under all of these generalizations of PD: the only Nash Equilibrium is one on which all players defect (the most), but (the most) cooperation leads to an outcome that is both Pareto dominant and consequently welfare dominant.

<sup>6</sup> If  $m$  individuals cooperated and  $n - m$  defected, then each cooperator will get a payoff of  $(m - 1) * R + (n - m) * S$  and each defector will get  $m * T + (n - m - 1) * P$ .

**Table 5** Left: Ordinary prisoner's dilemma. Right: PD with Punishment

	Distance	Not-Distance		Distance	Not-Distance
Distance	R, R	S, T	Distance	R, R	S, T-E
Not-Distance	T, S	P, P	Not-Distance	T-E, S	P-E, P-E

## Pandemic response as a social dilemma

Thinking of social distancing as an attractive (not risk-dominant but still a Nash equilibrium) policy depends on specific assumptions about the benefits of social distancing. A harsher conception is one in which social distancing is not even a Nash equilibrium. As in the prisoner's dilemma, the argument is one of dominance. If sufficiently many people are practicing social distancing, then any individual would benefit by defecting; from their point of view, they would avoid the costs of distancing while free-riding on the success overall. If very few people are practicing social distancing, then the epidemic will spread regardless, and any individual would be better by not incurring the cost of distancing. People who have a comparatively low risk of harm from infection might be prone to reason in this way. If so, then there is only one equilibrium: no one practices social distancing. This is because regardless of what others do, each individual would be at least as good by continuing to socialize. As in prisoner's dilemma, the world would be much better if everyone cooperates and the curve is flattened. But now the problem is much more difficult than merely *reaching* the good, stable equilibrium: the best outcome is unstable, and so we should not expect individuals to stay there.

This is a grim representation of the problem, but perhaps not an unrealistic one. The difference with conceiving it as a coordination problem is that here cooperation is not an equilibrium; most reasonable people, seeing that others have reason to defect, themselves have reason to defect too. Since large scale social distancing is not an equilibrium, we assume that accomplishing it will require further enforcement action by other individuals or by the state.

## Public signaling in social dilemmas

In strategic situations like the prisoner's dilemma, public signals alone cannot help us, at least if we understand signals as correlated strategy profiles relative to a mechanism. The only correlated equilibrium in a dilemma is one in which players defect. No matter the correlating device, defecting always beats cooperating, under any expectation of what others will do. Distancing is dominated by Not-Distancing under all expectations. In other words, agents are expected not to distance even when the option to distance is made salient to them.

One way of making the problem tractable is by considering punishments. Indeed, many public policy prescriptions—fines for failure to wear a mask, for example—are designed with this sort of scenario in mind. This could be done by considering a

Bayesian game in which players have a certain probability  $q$  of finding themselves in PD, and  $(1 - q)$  of finding themselves in a PD with punishment  $E$  for not distancing.

The transformation from the game in the left of Table 5 to the one on the right is not trivial. If  $E$  is sufficiently small, we would still be in PD and public signals will have no effect, as explained before. If  $E$  is larger, a PD can sometimes be transformed into a stag hunt.<sup>7</sup> In this case public signals would play a role. If  $E$  is sufficiently large, Distance would be the only equilibrium, and public signals will have no effect on this. Furthermore, an interesting question to consider is what is the relation between the degree of punishment and the strength of the signal that are jointly necessary and sufficient to secure widespread distancing. In other words, to study the interaction between the signal parameters  $p_i$  and the punishment parameters  $E$  and  $q$ . The more punishment  $E$  and the more expectation  $q$  of punishment, the less strength of signal will be necessary; and vice-versa. We leave this question for future research.

Disregarding scenarios where punishment is large (or small) enough to render public signaling ineffective, and noting that much depends on the the specific values of the parameters  $q$ ,  $T$ ,  $R$ ,  $P$ ,  $S$ ,  $E$ , our central point holds: a signal suggesting distance that does not provide enough expectation that others will distance will not be a correlated equilibrium.

### Lessons from the game theoretic analysis

Correlated strategies and mechanisms provide a clear formalization of the first and second-order messages delivered by epidemiological models that we introduced before. The former correspond to the suggestions provided by the model, the latter by what each agent expects others to do.

We argued that representing social distancing as a coordination problem or a social dilemma captures some of its most important features as a collective action problem. Furthermore, epidemiological models can be thought as playing a signaling role in these games.

Our game theoretic analysis revealed that in order to secure the socially optimal outcome, signals need not only give the right suggestions, but also do it in a way that individuals expect others to comply with that suggestion. But making sure that the right suggestion is the salient coordination point is sometimes in tension with the best modeling practices.

### Discussion

We have argued that publicly available epidemiological models play an important role in settling certain kinds of coordination problems and social dilemmas. We also argued that this role is obligatory: presentations of models can play this role

<sup>7</sup> If  $R = 3$ ,  $S = 0.5$ ,  $T = 3.5$ ,  $P = 2$ , and  $E = 1$

well or poorly, but it cannot be avoided. This creates a dilemma for modelers, as the requirements on a good scientific model can be in tension with the requirements for a strong social signal.

While the COVID-19 pandemic represents a particularly notable case of the coordination dilemma, we think that our result should apply more widely. Indeed, a version of the dilemma should arise in any case where public and private goods have the potential to compete, and in which scientific modeling about good outcomes provides a salient focal point for coordination. This will be the case in, for example, policy around mass vaccination, climate change, and other public health initiatives. Different cases may face the dilemma in more or less stark terms. All the more reason, then, to study the ways in which it arises, and the ways in which trade-offs between accuracy and coordination can be fixed.

Consider the case of vaccination. Under certain assumptions it may be reasonably modeled as a coordination game or a social dilemma; where individual incentives somehow conflict with the best public outcome (with almost everyone vaccinated thereby reducing the effects of the pathogen). Public information about the reliability and safety of the available vaccines will shape the public's response. As a salient example, consider the Australian response to AstraZeneca's COVID-19 vaccine (AZ). The European Medicines Agency found that AZ had a very small chance of causing a rare blood clotting disorder. Hence there was a trade-off between risk and benefit to different age groups, given that Australia also had access to a more limited stock of Pfizer vaccines. The Australian Technical Advisory Group on Immunisation (ATAGI) originally took a very cautious approach, recommending that AZ be given only to adults over 50. That advice changed several times in the subsequent months, in part as new outbreaks arose (changing the risk calculation), in part because of the challenges posed by more infectious strains of COVID-19, and in part because of conflicting communication by state and federal political leaders. The threshold was both raised and lowered, and the vaccine was finally cleared for anyone 18 and over.<sup>8</sup>

Each of these changes (let's suppose) was perfectly reasonable given the changing risk landscape, and so represented sound scientific advice. Yet one of the net results of these frequent changes appears to have been an rise in vaccine hesitancy, both overall, and with respect to AZ in particular. Indeed, at the time of writing, one of the primary reasons cited by over-70 patients for hesitancy is the desire to wait until another vaccine is available—despite the fact that over-70s have *always* been encouraged to take AZ if available.<sup>9</sup> This seems to be a real-life example of the coordination dilemma. ATAGI had good scientific reasons to alter its recommendation as the landscape of risk changed. Yet the changing risks seem to have undermined the public signal about the efficacy of AZ in particular, as well as a more general worry about vaccination as a whole.

---

<sup>8</sup> For a list of ATAGI's changing recommendations, see Australian Technical Advisory Group on Immunisation (2021).

<sup>9</sup> For Vaccine hesitancy over time see Essential Research (2021). For breakdown of reasons by age, see Australian Bureau of Statistics (2021).

The coordination dilemma has obvious policy implications. It may also be of interest to scholars working on what we called the standard dilemma. Work on the standard dilemma is partly descriptive, and is meant to track the actual practices of scientists and the trade-offs they perceive between virtues like simplicity and accuracy. We suggest that an appreciation of the coordination dilemma might shed further light on these first-order practices. One might expect researchers themselves to be aware of the fact that they face a coordination dilemma, and to re-weight these trade-offs accordingly when they know that their work might form a coordination point. Further work and modeling here might thus reveal a more complex set of interactions between first- and second-order considerations.

Once the coordination dilemma arises, *solving* it might depend on the particularities of the circumstances, and on solving a trade-off between scientific and broader social concerns. It is for this reason that we prefer to remain neutral on how to do it. Nevertheless, we argue, there are necessary conditions for the dilemma to emerge; it is not the case that all public health communication gives rise to it. We conclude by suggesting four background conditions that make the dilemma possible. By intervening in those conditions we might be able to ameliorate the tension between first- and second-order messages, and in some cases *dissolve* it.

First, for the dilemma to arise, there has to be *substantial uncertainty* about the appropriate course of action. This generates the possibility for competing first-order models, and so noise in the second-order signal. The early months of the COVID-19 pandemic were marked by widespread uncertainty about rates and routes of transmission, susceptibility of subpopulations, the effectiveness of various interventions, and so on. Conversely, the less disagreement there is among experts on details, at least in the broad sense, the less tension there is between clear messaging and effective science: this is, arguably, how topics like climate change have evolved.

For the dilemma to arise, models must be *available* and *accessible*. Availability of models means that the public can see the different modeling efforts—they are not pay-walled or treated as secret. The wide availability of COVID-19 models, including the use of preprint servers, meant that the public could see the divergence in different models as they played out. Accessibility means that the modeling is not completely esoteric, in the sense of (Goldman 2001): non-expert members of the public can look at the models and understand them well enough to know that they diverge. Unlike uncertainty, availability and accessibility seem more difficult to ameliorate: indeed, one might think that this is a necessary but unexpected side effect of the push towards more open science. Of course, there will inevitably be disparity of access, engagement, and scientific literacy. One future direction for this work might be to explore the effect of this disparity by introducing noise into the agents' received signals.

Fourth and finally, the dilemma only really gets off the ground when the coordination function of models is driven by the strategic considerations that we modeled with game theory. As is well-known in the literature, coordination problems can also be solved by (e.g.) social norms, strong institutions, or a shared sense of interpersonal trust. Under such conditions, one might expect the coordination function of models to play less of a role in overall coordination. Building up strong trust



in institutions, while never easy, might thus be a further way to lessen the sting of the modelers dilemma.

**Acknowledgements** We would like to thank the two anonymous reviewers as well as the audiences at the Australian National University, Sydney University, and Macquarie University Philosophy seminars for helping us improve the paper with their comments. Particular mentions should go to Jeremy Strasser, Rachael Brown, Katie Steele, and Atoosa Kasirzadeh for their observations. This paper was supported by Australian Research Council Grant DP190101507 (to CK) and funding from the Humanising Machine Intelligence Grand Challenge Project at ANU (to IOQ, SR, and CK).

## References

- Aumann RJ (1987) Correlated equilibrium as an expression of Bayesian rationality. *Econometrica* 55(1):1–18
- Australian Bureau of Statistics (2021) Household impacts of covid-19 survey. <https://www.abs.gov.au/statistics/people/people-and-communities/household-impacts-covid-19-survey/latest-release#covid-19-vaccination>
- Australian Technical Advisory Group on Immunisation (2021) Covid-19 vaccination statements. <https://www.health.gov.au/committees-and-groups/australian-technical-advisory-group-on-immunisation-atagi>
- Douglas H (2009) Science, policy, and the value-free ideal. University of Pittsburgh Press, Pittsburgh
- Essential Research (2021) Preference towards covid-19 vaccines. <https://essentialvision.com.au/preference-towards-covid-19-vaccines-4>
- Funk S, Gilad E, Watkins C, Jansen VA (2009) The spread of awareness and its impact on epidemic outbreaks. *Proc Natl Acad Sci* 106(16):6872–6877
- Funk S, Salathé M, Jansen VA (2010) Modelling the influence of human behaviour on the spread of infectious diseases: a review. *J Royal Soc Interface* 7(50):1247–1256
- Goldman AI (2001) Experts: which ones should you trust? *Philos Phenomenol Res* 63(1):85–110
- Goodin RE (1980) Manipulatory politics. Yale University Press, New Haven
- Harsanyi JC, Selten R (1988) A general theory of equilibrium selection in games. The MIT Press, Cambridge
- Lewis D (1969) Convention. Harvard University Press, Cambridge
- Lourenço J, Paton R, Ghafari M, Kraemer M, Thompson C, Simmonds P, Klenerman P, Gupta S (2020) Fundamental principles of epidemic spread highlight the immediate need for large-scale serological surveys to assess the stage of the SARS-CoV-2 epidemic. *MedRxiv*. <https://doi.org/10.1101/2020.03.24.20042291>
- Medvecky F, Leach J (2019) An ethics of science communication. Springer, Berlin
- Morsky B, Akçay E (2019) Evolution of social norms and correlated equilibria. *Proc Natl Acad Sci* 116(18):8834–8839. <https://doi.org/10.1073/pnas.1817095116>
- Reluga TC (2010) Game theory of social distancing in response to an epidemic. *PLoS Comput Biol* 6(5):e1000793
- Rudner R (1953) The scientist qua scientist makes value judgments. *Philos Sci* 20(1):1–6
- Schelling TC (1960) The strategy of conflict. Harvard University Press, Cambridge
- Skyrms B (2010) Signals: evolution, learning, and information. Oxford University Press, Oxford
- Skyrms B (2014) Evolution of the social contract. Cambridge University Press, Cambridge
- Vanderschraaf P (1995) Convention as correlated equilibrium. *Erkenntnis* 42(1):65–87

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.