Cassandra's Curse: A second tragedy of the commons

Philippe Colo (ETH Zurich)

Global Priorities Institute | September 2022

GPI Working Paper No .12-2022

GLOBAL PRIORITIES INSTITUTE

UNIVERSITY OF OXFORD

Cassandra's Curse:

A Second Tragedy of the Commons*

Philippe Colo [1]

[1]ETH Zurich

22nd Sept, 2022

Abstract

This paper studies why scientific forecasts regarding exceptional or rare events generally fail to trigger adequate public response. I consider a game of contribution to a public bad. Prior to the game, I assume contributors receive non-verifiable expert advice regarding uncertain damages. In addition, I assume that the expert cares only about social welfare. Under mild assumptions, I show that no information transmission can happen at equilibrium when the number of contributors is high or the severity of damages is low. Then, contributors ignore scientific reports and act solely upon their prior belief.

Keywords: Contribution to a public bad, Cheap talk, Climate change

*I thank Loic Berger, Valentina Bossetti, Adrien Fabre, Arnaud Goussbaille, Frédéric Koessler, Eugen Kovac, Hélène Ollivier, Marco Ottaviani, Jean-Marc Tallon and Stéphane Zuber for helpful discussions. I also thank seminar and conference participants at PSE (TOM, SRE), Bocconi, SAET 2019 and FAERE 2021 as well as Bocconi University and ETH Zurich for their hospitality. Financial support through ANR CHOp (ANR-17-CE26-0003), ANR ADE (ANR-18-ORAR-0005-01), ANR INDUCED (ANR-17-CE03-0008) and EUR PGSE is gratefully acknowledged.

In Greek mythology, Cassandra was a Trojan princess cursed by Apollo to utter true prophecies, but never to be believed. Many times, she used her gift to advise her fellow Trojans to be wary of the Greeks. Despite her warnings, her brother, prince Paris, brought the beautiful Helen back to Troy, starting a war that would be fatal to the city. Despite her warnings, Trojans accepted the famous wooden horse, gifted by the Greeks, that would end the conflict. From our perspective, it may seem obvious that the Trojans should have listened to Cassandra. But at the time her prophecies were made, were they really wrong to ignore her?

Consider the case of the arrival of Helen. Assume Cassandra cares about the good of all the Trojans. Paris, for his part, is also driven by the desire to fulfil his love for Helen. Because the presence of Helen poses a threat to the survival of the city, there is always a difference in motives between Cassandra and Paris. So, when Cassandra announces that she has foreseen that this arrival will trigger war with the Greeks, Paris will be reluctant to send Helen away. For him, even if Cassandra is right, his love for Helen is probably worth a war. Were Cassandra to anticipate Paris' reaction, she would have an incentive to effectively transform the truth and exaggerate the consequences of Helen's arrival. Because Paris is aware of that, if Cassandra chooses to tell the truth, he will assume she is exaggerating and ignore her warning.

In fact, Cassandra's lack of credibility goes beyond Paris alone. Most male Trojans are mesmerised by Helen's beauty and enjoy her presence in the city. Because of that, their interest is not aligned with the good of Troy as a whole—that is, including women and children who may be less keen on waging war for the beauty of a foreign queen. If Cassandra argues for the need to return Helen to Sparta, Trojan males will, for reasons similar to Paris, consider she is over-reacting. In the words of modern game theory, for Cassandra, convincing the Trojans of what she has foreseen is impossible at equilibrium.

In this paper I show that at equilibrium, Cassandras are unable not only to convey the truth to Trojans, but to convey any information at all. For perfectly rational Trojans, whatever Cassandra claims, she should be ignored when it comes to decision making. This result is all the more striking if one considers that in practice, Cassandra is telling the truth about what she knows. The strength of the theoretical analysis is to reveal that, because of the strategic setting, these truthful claims will be perceived as non-credible by her audience.

Sadly, climate scientists are present-day Cassandras. For decades, there has been scientific consensus on human responsibility for global warming. Considerable effort and international coordination has been devoted to raising awareness of the negative consequences of greenhouse gas (GHG) emissions on the environment, a task most famously embodied by the creation of the Intergovernmental Panel on Climate Change (IPCC) in 1988. Yet, multiple studies have reported that despite climate change being an important and widespread concern in the general population (Whitmarsh and Capstick, 2018), there is still strong resistance against concrete political measures such as carbon pricing.[1] Most of us act like the male Trojans in Cassandra's example. Our Helen is our carbon-intensive way of life. We enjoy it and stick to it just like Trojans enjoy the presence of Helen and keep her into their city. In doing so, we neglect the global effects of our decisions just as much as male Trojans did. Then, because of the strategic setting, even if climate scientists tell the truth about what they know—which they do—their claims are mostly ignored when it comes to concrete decisions.

The argument made here goes beyond the case of climate change. Cassandra's curse is rooted in the source of her knowledge: premonition. For her audience, the credibility of her claims depends entirely on the confidence they have in her. They cannot be verified through logic or perception. Cases like these, where knowledge depends solely on testimony, typically arise in the face of exceptional and rare events. Situations such as a new kind of global pandemic or the massive use of a vaccine relying on unprecedented technologies are good examples. Everett et al. (2021) argue that because the COVID-19 crisis was such an unexpected event, compliance with health recommendations has mostly been determined by the degree of trust populations have in their leaders. In the same vein, for a layman, mRNA vaccines are a new and unknown kind of technology. Their reliability cannot be assessed from the layman's past experience of standard vaccination. Hornsey et al. (2020) show that the sudden increase in distrust regarding vaccination we currently observe is best explained by a distrust in political and pharmaceutical actors.

[1] For instance, see Maestre-Andrés et al. (2019) or Douenne and Fabre (2020) for an empirical review of the determinants of the political acceptability of carbon pricing.

Of course, in the examples I have mentioned, counter-acting mechanisms that allow for some transmission of information are also at play. Scientific communication is a complex phenomena that cannot be summarised by a single model. Nevertheless, in order to better understand the specific effect I am interested in, I will consider the extreme case where scientific information is solely provided through non-certifiable communication. Arguably, in the 1980s, public knowledge of climate change was almost entirely expert-based. Scientific consensus was already formed on the topic, but its practical effects were mostly invisible for the general public. Therefor, in the remainder of this paper, I will focus on the case of climate change as a leading application.

Multiple empirical studies have been conducted to understand why scientific forecasts regarding climate change fail to trigger strong public willingness for its mitigation. Overall, it appears that public reluctance regarding climate regulation is less a problem of mistrust in the reality of climate change than linked to strategic considerations.

Free-riding on climate mitigation or the fear of others free-riding is generally the most common explanation identified by surveys (Fischer et al., 2011). As argued by Halady and Rao (2010), this aspect is made even worse by the presence of uncertainty regarding climate damage.

The complexity of climate science is a second element often put forward by empirical work. Sterman (2008) famously showed how hard it is for MIT students to draw a causal link between the multiple potential causes of climate change and their consequences. In the same vein, Lorenzoni and Pidgeon (2006) surveyed a large sample of citizens in Europe and the USA and showed that most people have limited knowledge about climate change and mostly relate to it through trust in experts. For most of us, the consequences of our actions on climate change is simply too complex to assess.

Lack of trust in scientists and information providers is a third aspect which can explain resistance to carbon-reducing measures. This may be due to normative reasons. Gabel et al. (2021) show that American conservatives display less confidence in science than liberals because scientists are believed to prioritise regulation over individual freedom. Ehret et al. (2018) and Van Boven et al. (2018) also provide evidence on how distrust for climate science can emerge from political partisanship and perceived normative misalignment.

In this paper, I combine these three aspects in a theoretical model to show why science, despite years of public communication over the upcoming climate catastrophe, struggles to trigger public desire for strong regulations.

First, to account for the free-riding aspect of climate change, I consider a game of contribution to a public bad, GHG emissions, where there is uncertainty over the damage resulting from global warming. Contributors benefit from emissions to the extent that carbon-intensive energies are needed to produce consumption goods, but expose themselves to potential climate damage. I will only assume that the marginal disutility caused by emissions is higher than its marginal benefit. I focus on the equilibrium level of emissions as a measure of the consumption level the public is prepared to accept.

Second, before choosing their emission levels, contributors receive advice from an expert regarding climate risk. However, the information she provides is not verifiable by contributors. This assumption captures the fact that climate science is too complex to be seen as verifiable information for the general public. Although scientific results regarding climate change are well established for scientists, for non-scientists the knowledge needed to understand the supporting evidence is simply too vast. The most striking examples are certainly *black box* prospective computer simulations, which are heavily relied on to estimate the effects of GHG on global temperatures. As pointed out by Pidgeon and Fischhoff (2011), even for scientists whose disciplines use observational methods, black box simulations are true to their name. They can hardly be considered as convincing evidence.

Third, to capture the normative differences between expert and layman, I assume that the former is known to be benevolent and to aim for the utilitarian, socially-optimal level of emissions. Because of the free-riding effect in the contribution game, under perfect information, there is always a difference between the expert's ideal level of emission and the equilibrium one. This leads to a difference in normative objectives between the expert, who aims for social welfare, and the contributors, who aim for their own individual good. In the context of this game, this difference naturally leads to a lack of trust in science.

Overall, the game I consider is a game of cheap-talk communication where an expert publicly addresses multiple contributors to a public bad. There are two main results. First, I show that when the number of contributors is high enough, whatever the severity of climate damage, no information transmission can happen at equilibrium. As a result, scientific information is ignored and emission levels are chosen solely based on the contributors' prior. Second, I show that even when the number of contributors is small, when the severity of climate damage is low enough, the previous result is extended and no information transmission can happen at equilibrium.

This result constitutes a second *tragedy of the commons*. If the expert was able to reveal her information, contributors would still over-emit, but social welfare would be higher. Were the expert to make do with the decentralised outcome, she would be able to credibly reveal her information. Then, contributors' inefficiency would lead to a recognised fall in welfare, compared to the first-best situation, which has been designated as the tragedy of the commons. But the expert's aim for efficiency makes her a Cassandra and prevents information transmission. As a result, contributors must choose under full uncertainty, which leads to a second fall in welfare compared to the first-best situation.

Related literature— The choice of emission levels is in the tradition of the canonical public good game of Bergstrom et al. (1986). An important characteristic of games of contribution to a public good (or bad), when uncertainty is introduced, is that more information does not always increase social welfare.[2] In the context of games of information transmission, this observation is the driving force to understand equilibrium behaviour. Asheim (2010) was the first to study information transmission in the context of public good games. In his model, a benevolent environmental agency can certify information about climate risk to the contributors. He shows that full revelation is not always optimal for the environmental agency. In the same vein, Kakeu and Johnson (2018) study information exchange between privately-informed countries in a model of transnational pollution. Information transmission is certifiable but costly. They show that sharing information is only incentive-compatible under sufficiently precise private information. Slechten (2020) finds similar results regarding the role of (certifiable) information-sharing regarding abatement costs prior to an

[2] See Angeletos and Pavan (2007) for a more systematic review of games where information can have a negative value.

environmental agreement. Only when uncertainty is high enough will information transmission through the use of a mechanism be desirable for participants. The results of the present paper are in line with the general idea conveyed by this literature: because information transmission does not always have a positive value in games of contribution to a public good, at least some communication might not occur at equilibrium.

This paper is also the first to introduce non-certifiable information in the context of climate policies. It thus relates to the literature on non-certifiable strategic communication, starting with Crawford and Sobel (1982)'s cheap-talk model with one sender and one receiver. Goltsman and Pavlov (2008) study cheap talk with public and private communication with multiple receivers. However, unlike in my paper, receivers' actions do not affect each other. Interaction among receivers after the communication phase is a novelty of this paper, although Galeotti et al. (2013) had studied cheap-talk communication in the context of networks, where multiple imperfectly-informed senders are also decision-makers and can influence each other by both their actions and messages. A second theoretical novelty introduced by this paper is to provide a characterisation result in the context of a cheap-talk game where the misalignment between parties is not linear, as generally assumed in applications.

Section 1 introduces the base model. Section 2 solves the game of contribution to a public bad for any given message of the informed party. Section 3 establishes when no information can be communicated by the sender at equilibrium and discusses the welfare implications. Section 4 discusses the main implications of the results. The appendix argues that the results are robust to a more general choice of prior belief, utility functions and to heterogeneous preferences among contributors.

1 Setup

I consider a game between a scientific authority acting as a sender S (she), and N receivers R_i (he) who have to choose a level of GHG emissions $e_i \geq 0$. The set of possible actions for each receiver is thus $\mathcal{A} = \mathbb{R}^+$. Receiver i benefits from his emissions through consumption but suffers from the overall emission level. In addition, for a given level of total GHG emissions, there is uncertainty about the severity of damage suffered by the receivers. Let $\Omega = [a, b]$, where $b > a > 0$, represent the possible warming potentials of GHG emissions,[3] where climate damage is increasing with ω. I will refer to them as the states. The scientific authority S learns the state from Nature, but cannot certify her information. She can send to the receivers a costless message $m \in \mathcal{M}$, where $\mathcal{M}$ is a non-degenerate interval of $\mathbb{R}$, indicating what the state might be. The timing of the game is as follows:

1. Nature draws the state $\widetilde{\omega}$ according to a uniform distribution over Ω of density g.
2. S is privately informed of $\widetilde{\omega}$, which becomes her type.
3. **Communication stage:** the sender sends a message to the receivers regarding her type.
4. **Emission stage:** the receivers simultaneously choose a level of emissions e_i.

Receiver i's utility function will be as follows:

$$u_i(e_i, e_{-i}, \omega) = e_i - \frac{\omega}{\beta} \sum_{i=1}^{N} e_i^{\beta}$$

where $\beta > 1$ parametrises the severity of climate damage. Thus, individual GHG emissions benefit the receivers linearly, but their overall level negatively impacts them in a convex way. In addition, the higher ω is, the more severe the consequences of global warming and the greater the cost of emissions. Thus, the receivers' choice of emission level is the result of a trade-off between economic growth and potential damage caused by global warming. Whatever $\omega \in \Omega$ and $e_1, ..., e_N \in \mathbb{R}^+$ u_i is concave. Because $a > 0$, emissions will result in damage even when $\omega = a$ is strictly positive, expressing the fact that global warming is not avoidable in any state. Notice also that utility

[3]This is often referred to as the climate sensitivity parameter in climate science. Uncertainty over it has been extensively discussed (Meinshausen et al., 2009).

functions are single-peaked. As a result, for a given state, there is a single optimal emission level $e_i(\omega)$. A higher emission level $e > e_i(\omega)$ is not optimal for i because it might cause too much climate damage. A lower emission level $e < e_i(\omega)$ is neither optimal for i as it implies to reduces economic growth by too much. Finally, notice that receiver i does not take into account the impact of his emissions on other receivers. The sender, on the contrary, takes into account the externalities in receivers' behaviour and seeks to maximise social welfare. Her ex-post utility is:

$$
\begin{aligned}
u_S(e_1, ..., e_N, \omega) &= \sum_{j=1}^{N} u_j(e_j, e_{-j}, \omega) \\
&= \sum_{j=1}^{N} e_j - \frac{N\omega}{\beta} \sum_{j=1}^{N} e_j^{\beta}
\end{aligned}
$$

A strategy for S is $\sigma : \Omega \to \mathcal{M}$, which consists in transmitting a message m to the receivers regarding her private information. A strategy for a receiver consists in choosing an emission level as a function of message m. In the following, I will focus on perfect Bayesian equilibria. An equilibrium consists in a signalling strategy $\sigma(\omega)$ and an action rule for each receiver $y_1(m), ..., y_N(m)$ such that:

1. S chooses a strategy σ such that for all $m \in \mathcal{M}$:

$$u_S(y_1(\sigma(\omega)), ..., y_N(\sigma(\omega)), \omega)) \geq u_S(y_1(m), ..., y_N(m), \omega)$$

2. Having received an equilibrium public message $m \in supp(\sigma)$, R_i updates his prior using Bayes' rule such that:

$$g(\omega|m) = \begin{cases} \frac{g(\omega)}{\int_{\sigma^{-1}(m)} g(\omega)d\omega} \text{ if } \omega \in \sigma^{-1}(m) \\ 0 \text{ if not} \end{cases}$$

and chooses action $y_i(m)$, such that for all $e \in \mathcal{A}$:

$$\mathbb{E}(u_i(y_i(m), y_{-i}(m), \omega)|m) \geq \mathbb{E}(u_i(e, y_{-i}(m), \omega)|m)$$

where $\mathbb{E}(u_i(e_i, e_{-i}, \omega)|m) = \int_{\omega\in\Omega} g(\omega|m)u_i(e_i, e_{-i}, \omega)d\omega$. Any message m such that $m \notin supp(\sigma)$ is interpreted as some equilibrium message $m^* \in supp(\sigma)$.

2 Emission stage

I start by focusing on the emission stage. Consider any message $m \in \mathcal{M}$ and for any $i \in 1, ...N$, set $e_i(m)$ the solution to the maximisation problem:

$$\max_{e_i \in \mathbb{R}^+} \mathbb{E}(u_i(e_i, e_{-i}(m)|m)$$

$e_i(m)$ is thus the equilibrium level of emissions of receiver i having received message m. The first order condition gives that the equilibrium level of emission of receiver i must be such that:

$$e_i(m) = \frac{1}{\mathbb{E}(\omega|m)^{\frac{1}{\beta-1}}}$$

As emission levels are symmetric, in the following I will focus on the total level of emissions. Thus, for any given equilibrium message m, the equilibrium total emission level of the receivers is:

$$t(m) = \frac{N}{\mathbb{E}(\omega|m)^{\frac{1}{\beta-1}}} \tag{1}$$

Similarly, for any $i \in 1, ...N$, set $e_i^W(m)$ the solution to the maximisation problem:

$$\max_{e_i^W \in \mathbb{R}^+} \sum_{i=1}^{N} \mathbb{E}(u_i(e_i^W, e_{-i}^W(m), \omega)|m)$$

$e_i^W(m)$ is thus the socially optimal level of emissions of receiver i having received message m. As before, I will restrict attention to $t^W(m)$, the socially optimal total level of emissions. The first order condition gives that the socially optimal total level of emissions of the receivers must be such that:

$$t^W(m) = \sum_{i=1}^{N} e_i^W(m) = \frac{N}{(N\mathbb{E}(\omega|m))^{\frac{1}{\beta-1}}}$$

Thus, notice that whatever the message $m \in \mathcal{M}$, we have that:

$$t^W(m) = \frac{1}{N^{\frac{1}{\beta-1}}} t(m) \tag{2}$$

When $N = 1$, sender and receivers have exactly the same utility function and aim for the same total level of emissions. However, when $N > 1$, emission levels are always higher in the non-cooperative equilibrium than what would be socially optimal: $t^W(m) < t(m)$. The greater the number of inefficient receivers N, the greater that difference. In the following, I will denote $t^W(B)$ and $t(B)$, for B a subset of Ω, the socially optimal and decentralised levels of emission under the belief that $\widetilde{\omega} \in B$. To avoid confusion, I will assume that $\mathcal{M}$ and Ω are disjoint. By a slight abuse of notation, I will write $t^W(\omega)$ and $t(\omega)$ when B is a singleton.

3 Communication stage

Failure of communication in large societies— I now turn to the communication stage and state the paper's first main result. It provides a sufficient condition for no information transmission to be possible at equilibrium.

Theorem 1. *For $\beta > 1$, if $N \geq \frac{1}{2}(1+\frac{b}{a})$, there is no equilibrium where information transmission happens. For any $\omega \in \Omega$ and any message $m \in \mathcal{M}$:*

$$g(\omega|m) = g(\omega)$$

When the damage function is strictly convex and the number of receivers is large enough, no information transmission is possible at equilibrium. When $N \geq \frac{1}{2}(1+\frac{b}{a})$, I say the sender faces a *large society*. Thus, Theorem 1 states that in large societies, whatever the state that the sender learns, no message has the power to change the receivers' belief. Notice that what defines a society as large depends on the support for receiver's belief. If there is a strictly positive probability that damage is low (that is, if a is close to 0), even with two receivers, information transmission is impossible. On the contrary, if the support for receiver's belief is wide (that is, if b is much greater than a), the number of receivers must be larger for information transmission to be impossible.

In other words, intuitively, what Theorem 1 shows is that when uncertainty is limited (a close to b), it is better for receivers to ignore information coming from an expert who, although perfectly informed, is known to have misaligned preferences. To the contrary, when uncertainty is large (b is far away from a), the trade-off resolves the other way around: information provided by a biased expert is preferable to purely prior-based decision.

To see where this result comes from let us first focus on the nature of the strategic interaction between the sender and the receivers in the communication stage. The sender and the receivers are engaged in cheap-talk communication. The sender cares only about the total emission level of the receivers. As equation (2) shows, under perfect information, as long as $N \geq 2$, there is always a difference between what the sender would like to see as the total emission level of the receivers and what they actually do. Combined with the fact that S's utility function is continuous and single-peaked, this observation suffices to show the following result.

Proposition 1. *There can only be a finite number of equilibria in the communication stage. All of them are partitional.*

Proof of Proposition 1:

As the socially optimal emission level of each receiver is the same, one can equate the sender's utility function with $u_S(t,\omega) = t - \frac{\omega}{\beta N^{\beta-2}} t^{\beta}$ where t is the total emission level of receivers.

The proof is structured as follows: first, I show that the number of aggregate emission levels of the receivers induced at equilibrium is finite (lemma 1). Then, I prove that the set of types which get the same equilibrium outcome must form an interval. The continuity and the strict monotonicity of the sender's preferences close the argument.

Lemma 1. *There exists $\epsilon > 0$ such that if u and v are actions induced in equilibrium, $|u - v| \geq \epsilon$. Furthermore, the set of aggregate emission levels induced in equilibrium is finite.*

Proof of Lemma 1

I say that action u is induced by an S-type ω if it is a best response to a given equilibrium message m: $u \in \{t(\omega)|\omega \in \sigma^{-1}(m)\}$. Let Y be the set of all actions induced by some S-type ω. First, notice that if ω induces $\bar{t}$, it must be that $u_S(\bar{t},\omega) = \max_{t \in Y} u_S(t,\omega)$. Since u_S is strictly concave, it can take on a given value for at most two values of t. Thus, ω can induce no more than two levels of aggregate emission level of the receivers in equilibrium.

Let u and v be two levels of aggregate emissions induced in equilibrium, $u < v$. Define Θ_u the set of S types who induce u and Θ_v the set of S types who induce v. Take $\omega \in \Theta_u$ and $\omega' \in \Theta_v$. By definition, ω reveals a weak preference for u over v and ω' reveals a weak preference for v over u, that is:

$$\begin{cases} u_S(u, \omega) \geq u_S(v, \omega) \\ u_S(v, \omega') \geq u_S(u, \omega') \end{cases}$$

Thus, by continuity of $\omega \to u_S(u, \omega) - u_S(v, \omega)$, there is $\hat{\omega} \in [\omega, \omega']$ such that $u_S(u, \hat{\omega}) = u_S(v, \hat{\omega})$. Since u_S is strictly concave, we have that:

$$u < t^W(\hat{\omega}) < v$$

Then, notice that since $\frac{\partial^2 u_S(t,\omega)}{\partial t \partial \omega} > 0$, it must be that all types that induce u are below $\hat{\omega}$. Similarly, it must be that all types that induce v are above $\hat{\omega}$. That is:

$$\forall \omega \in \Theta_u, \omega \leq \hat{\omega}$$

$$\forall \omega \in \Theta_v, \omega \geq \hat{\omega}$$

Given that u_i, for $i \in 1, ..., N$, verify the assumptions of Crawford and Sobel (1982), the sum of optimal action of the receivers, given that $\omega \in \Theta_u$ is below the optimal action when the type is $\hat{\omega}$. Similarly, the sum of optimal actions of the receivers, given that $\omega \in \Theta_v$ is above the optimal action when the type is $\hat{\omega}$. That is:

$$\begin{cases} t(\Theta_u) \leq t(\hat{\omega}) \\ t(\Theta_v) \geq t(\hat{\omega}) \end{cases}$$
$$\iff u \leq t(\hat{\omega}) \leq v$$

However, as $t(\omega) \neq t^W(\omega)$ for all $\omega \in \Omega$, there is $\epsilon > 0$ such that $|t(\omega) - t^W(\omega)| \geq \epsilon$ for all $\omega \in \Omega$. It follows that $|u - v| \geq \epsilon$.

For any belief $B \subset \Omega$, the sum of optimal action of the receivers is in $\left[\frac{N}{b^{\frac{1}{\beta-1}}}, \frac{N}{a^{\frac{1}{\beta-1}}}\right]$. Thus, the set of actions induced in equilibrium is bounded and at least ϵ away from one another, which completes the proof.

□

Notice also that because u_S verifies all the requirements of Crawford and Sobel (1982), in every equilibrium of the game, if t is a level of aggregate emissions induced by type ω and by type $\omega^{''}$ for some $\omega < \omega^{''}$, then t is also induced by $\omega^{'} \in (\omega, \omega^{''})$.

By Lemma 1 there is a finite number of outcomes induced in equilibrium. The continuity of $t^W(\omega)$ gives that there is a type of the sender who is indifferent between any pair of outcomes induced in equilibrium and the monotony of $t^W(\omega)$ implies there are only a finite number of types who are indifferent between any pair of outcomes. Hence, the point made just above implies that there is a partitioning of Ω in a finite number of cells where every cell induces a given level of aggregate emissions at equilibrium. This implies that any equilibrium is a partition equilibrium.

□

In order to understand this result, let us first define what a partitional equilibrium is:

Definition 1. *Consider* $\{\omega_0, ..., \omega_q\} \subseteq [a, b]$ *such that:*

- *$a = \omega_0 < ... < \omega_q = b$ where ω_k, for $0 \leq k \leq q$, is called the k-th cut-off.*
- *$\cup_{k=1}^{q}[\omega_{k-1}, \omega_k] = [a, b]$, where $[\omega_{k-1}, \omega_k)$, for $1 \leq k < q-1$, is called the k-th cell and $[\omega_{q-1}, 1]$ the q-th cell.*

A q-cut-off partition equilibrium is an equilibrium of the game where the signalling strategy of S is uniform on every cell. That is, for $\omega \in [\omega_{k-1}, \omega_k)$, $\sigma^*(\omega) = m_k$, for $1 \leq k \leq q-1$ and for $\omega \in [\omega_{q-1}, 1]$, $\sigma^*(\omega) = m_{q-1}$.

Figure 1 illustrates the structure of a partitional signalling strategy. In this example, any type of sender in $[a, \omega_1]$ sends message $m_0 \in \mathcal{M}$, any type of sender in $[\omega_1, \omega_2]$ sends message $m_1 \in \mathcal{M}$, and any type of sender in $[\omega_2, b]$ sends message $m_2 \in \mathcal{M}$. In principal, multiple equilibria can exist, each of which is characterised by its cut-off types. There is always at least a 1-cut-off equilibrium where all types send the same message (and the cut-off then is $\omega_1 = b$). Because this equilibrium is uninformative, this equilibrium is called the babbling equilibrium.

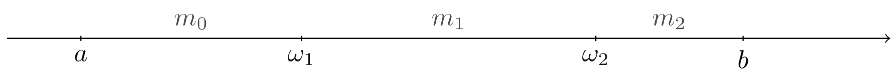

Figure 1: Partitional signalling strategy

Thus, a consequence of proposition 1 is that information transmission is always imprecise. As a result, the sender can never fully reveal her type, although some information transmission is still possible through the partitioning of Ω. For this to be possible, there must be adjacent intervals of types (cells) separated by a cut-off type where all types on one side prefer separating themselves from the types on the other side. Figure 2 illustrates how this may happen. An essential assumption for this to be possible is again that the utility function of the sender is single-peaked. Therefore, there can be a type (ω_k in figure 2) who is exactly indifferent between signalling herself as being part of the cell below her (by sending message m_{k-1} and triggering $t(m_{k-1})$ emissions) or above her (by sending message m_k and triggering $t(m_k)$ emissions). In addition, because optimal actions are strictly increasing with the type, it must be that all types above this cut-off strictly prefer distinguishing themselves from the types below it and vice versa.

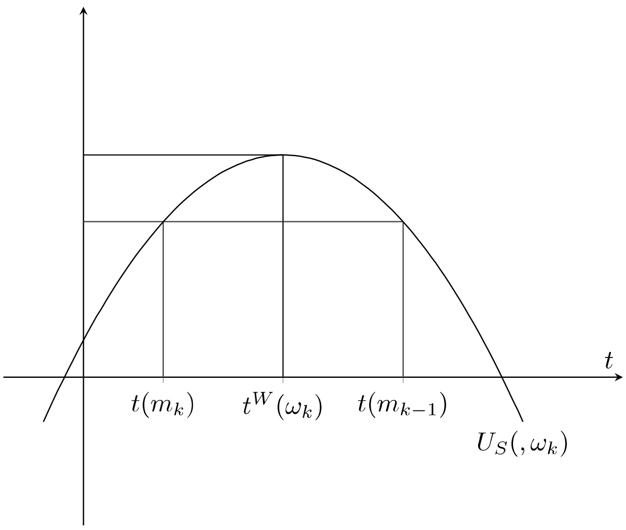

Figure 2: Identifying cut-off ω_k for $\beta = 2$

Thus, for there to be an informative equilibrium strategy of the sender, there must be $\omega_0 \in (a, b)$ such that a sender of type ω_0 is indifferent between signalling herself as being part of the cell above or below her. For this to be the case, because the sender's preferences are single-peaked, a necessary condition is that:

$$t([\omega_0, b]) \leq t^W(\omega_0)$$

That is, if the sender were to credibly signal that the state is above a given threshold ω_0, it must at least be that the resulting emissions are below the socially optimal level at that threshold. However, this condition is equivalent to:

$$\frac{N}{(\frac{\omega_0+b}{2})^{\frac{1}{\beta-1}}} \leq \frac{N}{(N\omega_0)^{\frac{1}{\beta-1}}}$$
$$\iff \omega_0 \leq \frac{b}{2N-1}$$

Yet, it must also be that $\omega_0 \in (a, b)$. As a result, a sufficient condition for ω_0 not to be a cut-off is:

$$\begin{aligned} \frac{b}{2N-1} &\leq a \\ \iff N &\geq \frac{1}{2}(1+\frac{b}{a}) \end{aligned}$$

which gives Theorem 1. Notice that the severity of climate damage β played no role in the analysis, apart from the assumption that $\beta > 1$.

Failure of communication for small societies and low damage— In large societies, no information transmission is possible at equilibrium. In the following, I show that even when the number of receivers is low enough, information transmission can still be impossible.

Theorem 2. *When $\beta \leq 2$, for $N \geq 2$, there is no equilibrium where information transmission happens. For any $\omega \in \Omega$ and any message $m \in \mathcal{M}$:*

$$g(\omega|m) = g(\omega)$$

Proof of Theorem 2:

First consider the case where $\beta = 2$. Then:

$$\begin{aligned} t^W(\omega_0) &= \frac{t(m^-) + t(m^+)}{2} \\ \iff \quad \frac{1}{4N\omega_0} &= \frac{1}{\omega_0 + a} + \frac{1}{b + \omega_0} \end{aligned}$$

Define:

$$f : \omega_0 \to \frac{1}{\omega_0 + a} + \frac{1}{b + \omega_0} - \frac{1}{4N\omega_0}$$

The situation where there are only two receivers, $N = 2$, is the situation where potential communication is the easiest to sustain. Then, the only positive candidate cut-off type is:

$$\begin{aligned} & f(\omega_0) = 0 \\ \iff \quad & \omega_0 = \frac{1}{30}(\sqrt{49a^2 + 158ab + 49b^2} - 7a - 7b) \end{aligned}$$

Yet, $\frac{1}{30}(\sqrt{49a^2 + 158ab + 49b^2} - 7a - 7b) < a$, for any $b > a > 0$. Notice that for $\omega_0 \in (a, b)$, f is strictly positive.

Because the sub-game played in the communication stage is a special case of the game studied by Crawford and Sobel (1982), by applying their Theorem 1 it follows that if there is no 2 cut-off equilibrium, the only equilibrium is the babbling one.

Now consider the case where $\beta < 2$. The marginal disutility of under-emitting is greater than that of over-emitting. Intuitively, the optimal action of the sender at a cut-off type $(t^W(\omega_0))$ has to be closer to the one induced by revealing her message above that threshold $(t(m^+))$ than below it $(t(m^-))$.

Thus, when the sender is indifferent between $t(m^-)$ and $t(m^+)$, it must be that there is $\alpha(\beta) \in (0, 1)$ such that:

$$t^W(\omega_0) = \alpha(\beta)t(m^-) + (1 - \alpha(\beta))t(m^+)$$

Lemma 2 establishes that $\alpha(\beta)$ is an decreasing function of β.

Lemma 2. $\alpha(\beta)$ *is an decreasing function of* β

Proof of Lemma 2:

For any given state ω, take $A^- < t^W(\omega) < A^+$ such that $u_S(A^-, \omega) = u_S(A^+, \omega)$.

For the utility level $u_S(A^-, \omega) = u_S(A^+, \omega)$, we have that:

$$\alpha(\beta) = \frac{t^W - A^-}{A^+ - A^-}$$

Take the tangent equation between A^- and t^W that gives that $u_S(A^-, \omega) = u_S(t^W, \omega) - \frac{\partial u_S(A^-,\omega)}{\partial t}(t^W - A^-)$. We then get that:

$$\alpha(\beta) = \frac{1}{A^+ - A^-} \times \left(u_S(t^W, \omega) - u_S(A^-, \omega)\right) \times \frac{1}{\frac{\partial u_S(A^-,\omega)}{\partial t}}$$

$\frac{1}{A^+-A^-}$ is constante in β. Because u_S is concave, $\frac{\partial u_S(A^-,\omega)}{\partial t}$ is decreasing with β and $u_S(t^W, \omega) - u_S(A^-, \omega)$ is decreasing in β. It follows that $\frac{1}{\frac{\partial u_S(A^-,\omega)}{\partial t}}$ is decreasing in β. As a result $\alpha(\beta)$ is decreasing in β.

□

As $\beta = 2$, $\alpha(\beta) = \frac{1}{2}$, Lemma 2 gives that $\beta < 2 \Rightarrow \alpha(\beta) \geq \frac{1}{2}$. Thus to prove that there can not be a 3 cut-off equilibrium when $\beta < 2$ it is sufficient to show that for any $\omega_0 \in (a, b)$:

$$\begin{aligned} & t^W(\omega_0) < \frac{t(m^-) + t(m^+)}{2} \\ \iff & \Big(\frac{1}{2N\omega_0}\Big)^{\frac{1}{\beta-1}} < \frac{1}{2}\Big[\Big(\frac{1}{\omega_0 + a}\Big)^{\frac{1}{\beta-1}} + \Big(\frac{1}{b+\omega_0}\Big)^{\frac{1}{\beta-1}}\Big] \\ \iff & \frac{1}{2N\omega_0} < \Big(\frac{1}{2}\Big[\Big(\frac{1}{\omega_0 + a}\Big)^{\frac{1}{\beta-1}} + \Big(\frac{1}{b+\omega_0}\Big)^{\frac{1}{\beta-1}}\Big]\Big)^{\beta-1} \end{aligned}$$

Yet, when $1 < \beta < 2$, $t \to t^{\beta-1}$ is a strictly concave and increasing function, we also have that:

$$\Big(\frac{1}{2}\Big[\Big(\frac{1}{\omega_0 + a}\Big)^{\frac{1}{\beta-1}} + \Big(\frac{1}{b+\omega_0}\Big)^{\frac{1}{\beta-1}}\Big]\Big)^{\beta-1} \geq \frac{1}{2}\Big[\frac{1}{\omega_0 + a} + \frac{1}{b+\omega_0}\Big]$$

Yet, the case $\beta = 2$ gave us that for $\omega_0 \in (a, b)$:

$$\frac{1}{4\omega_0} < \frac{1}{2}\Big[\frac{1}{\omega_0 + a} + \frac{1}{b+\omega_0}\Big]$$

It follows that necessarily, for $\omega_0 \in (a, b)$:

$$t^W(\omega_0) < \frac{t(m^-) + t(m^+)}{2}$$

which concludes the proof.

□

Theorem 2 gives the same result as Theorem 1 but for small damages ($\beta \leq 2$) and any number of receivers. To understand this result, first, consider the case where $\beta = 2$. For there to be at least a two cut-off equilibrium strategy, there must be $\omega_0 \in (a, b)$ such that a sender of type ω_0 is indifferent between signalling herself as being part of the cell below her (by sending message m^- and triggering $t(m^-)$ emissions) or above her (by sending message m^+ and triggering $t(m^+)$ emissions). When $\beta = 2$, for the sender to be indifferent between two aggregate emission levels, it must be that they are at equal distance from her optimal action. Formally, there must be $\omega_0 \in (a, b)$ such that:

$$t^W(\omega_0) = \frac{t(m^-) + t(m^+)}{2}$$

However, the resolution of the above equation in the proof of Theorem 2 gives that for any $\omega_0 \in (a, b)$, $t^W(\omega_0) < \frac{t(m^-)+t(m^+)}{2}$. Intuitively, whatever the state, quadratic damage is too low to trigger emission levels that are low enough to meet the sender's credibility constraint. Thus, when $\beta = 2$, there is no two cut-off equilibrium. In the case where $\beta < 2$, damage is even lower than when $\beta = 2$. As a result, it is not surprising that an informative signalling strategy does not, in any state, trigger emission levels that are low enough to meet the sender's credibility constraint.

Risk prudence— By varying the level of the severity of climate damage β as we have been doing until here, it should be noted that we are implicitly changing the receivers' level of risk prudence. To see this, notice that:

$$\frac{\partial^3 u_i(e_i, e_{-i}, \omega)}{\partial e_i^3} = -\omega(\beta - 1)(\beta - 2)e_i^{\beta-3}$$

which leads to the following remark:

Remark 1. *Receivers display risk prudence if and only if* $1 < \beta < 2$.

In my setting, risk prudence characterises the fact that receivers prefer greater risk in optimistic states (closer to a) than in pessimistic ones (closer to b). Compared to the case where $\beta \geq 2$, risk prudence will thus make receivers increase their level of emissions under message m^- but reduce it under message m^+. A second reading of Theorem 2 is thus that risk prudence necessarily leads to failure of communication, whatever the number of contributors.

Welfare analysis— It is well known that under perfect information, because receivers fail to internalise the consequences of their actions on others, the overall level of emissions is inefficient at equilibrium. This situation corresponds to the case where the sender makes do with the decentralised outcome instead of aiming for social welfare. That is, if for any $\omega \in \Omega$, $t^W(\omega) = t(\omega)$, the expected social welfare will be:

$$W_R(\beta) \quad = \quad \int_a^b t(\omega) - \frac{\omega N}{\beta}\Big(\frac{t(\omega)}{N}\Big)^\beta d\omega$$

Consider the case where damage is low ($\beta \leq 2$) or society is large ($N \geq \frac{1}{2}(1 + \frac{b}{a})$). When the sender aims for social welfare and, for any $\omega \in \Omega$, $t^W(\omega) < t(\omega)$, the expected social welfare will be:

$$W(\beta) \quad = \quad \int_a^b \int_a^b t(\omega^{'})d\omega^{'} - \frac{\omega N}{\beta}\Big(\frac{\int_a^b t(\omega^{'})d\omega^{'}}{N}\Big)^\beta d\omega$$

However, because for $\beta > 1$, $t \to t^\beta$ is a strictly convex function, we also have that:

$$\int_a^b t(\omega^{'})^{\beta} d\omega^{'} > \Big(\int_a^b t(\omega^{'}) d\omega^{'} \Big)^{\beta}$$

It immediately follows that for any $\beta > 1$, $W(\beta) < W_R(\beta)$. That is to say, because of uncertainty, the expected welfare when the sender aims for the socially optimal level of emissions is lower than when she makes do with the decentralised level. We thus get the following result:

Proposition 2. *When $\beta \leq 2$ or $N \geq \frac{1}{2}(1 + \frac{b}{a})$, the expected social welfare of the game is lower when the expert is benevolent than when she aims for the decentralised outcome.*

Arguably, this situation can be called a second tragedy of the commons. It is the result of a difference in incentives for the sender between the ex-ante perspective and the interim one. Before learning the state (ex-ante), the sender knows she is better off under full revelation than under cheap talk. If she could commit to conveying her information at that moment, she would. Yet, once she learns her type (interim), the difference between the socially optimal level of emissions and the equilibrium level is such that even the most optimistic types have no incentive to separate themselves from the others. In other words, with no other argument in support of her good faith than her incentives, no informed sender is capable of conveying information, leaving the receivers under full uncertainty. The convexity of the damage function then implies that the latter situation is strictly worse than the former for social welfare.

4 Discussion

What do we learn from these results? As a theoretical model, the present work's main contribution is to offer insights on the weight of our assumptions when we think of Cassandra-like situations.

Welfarism— A crucial aspect to discuss is the assumed objective function of the sender. Here, the sender aims for the utilitarian social welfare. In function of the application, this can certainly be challenged. Yet, there are good reasons to think that, in most cases, there is a misalignment between the optimal action of the sender and the decentralised outcome of the receivers' game. Climate scientists or health authorities generally have more holistic motivations—although not necessarily strictly welfarist—than what would come out from a simple *laissez faire*. Therefor, this paper's main message is twofold. First, it has a positive dimension: when a sender's motivation lean towards welfarism, her messages may be inaudible for receivers engaged in a game of contribution to a public bad. Second, it also has a normative dimension: when evaluating the virtues of the welfarist approach in the concrete cases we mentioned here, it is essential to account for the (negative) strategic effects we have put forward.

Single sender— On many instances, there is more than one expert communicating with the public on the state of science. Often, they are not perfectly informed and therefor they may disagree. The theoretical literature on multiple sender cheap-talk communication games has mixed results. Battaglini (2002) and Battaglini et al. (2004) show that simultaneous communication generally lead to full revelation while Krishna and Morgan (2001) show that under sequential communication information transmission is still partial. The present model applies when the informed parties are perceived as sufficiently unified by the decision makers to be able to coordinate on their public policy communication. As argued by Tol (2011), the IPCC's monopolistic situation on climate knowledge could enable this coordination. On the COVID-19 vaccination example, the World Health Organization may play that role. What this paper shows is that a strategic counterpart of this coordination possibility is a potential failure in information transmission.

Damage functions— In the climate example, the range of possible climate sensitivity parameters is broad and the severity of potential climate damage is high. Following Theorem 1, both aspects favour the possibility of some information transmission to a small group of contributors—for instance, a small group of political leaders. However, when communicating with the general public, information transmission would still fail. Conversely, in the COVID-19 vaccination example, if the vaccination rate is too low, negative consequences are relatively clear to predict. In other words, the spread of possible consequences is not too large. In addition, compared to the climate case, the absolute value of the marginal cost of a low vaccination rate is also closer to its marginal benefit. Following Theorem 2, the two latter observations favour situations where information transmission is impossible, even to a small group.

Appendix

How general are the results presented until here? The primary goal of the analysis was to present a coherent framework consistent with empirical observations and providing an insight to why scientific forecasts regarding, for instance, climate change are unable to trigger its mitigation. Therefor, I did not force the model to be more complicated than is necessary. However, In the following, I discuss how some of the assumptions I have made can be relaxed and how that would affect the result.

The role of the prior— Clearly, the assumption of a uniform prior over Ω has been made for computational convenience. Although it has an influence on the main result, I argue this influence is limited. To see this, consider two extreme cases: the one of an extremely optimistic prior of full support but where almost all the probability mass concentrates towards the lower bound of Ω, and the symmetric case of an extremely pessimistic prior.[4]

Intuitively, the optimistic case is even worse for communication than the one with uniform prior. This is because, for any given message, optimal actions of the receivers are shifted to the right, compared with the uniform prior case, while, for any state, the optimal action for the sender is lower than the equilibrium one. As before, consider the 2 cut-off partitional strategy σ where the interior cut-off is ω_0. In the limit of the optimistic prior, we have that m^+ induces action $t(\omega_0)$. Yet we have that:

$$t(\omega_0) > t^W(\omega_0)$$

making it impossible for ω_0 to be a cut-off type.

On the contrary, the pessimistic case should favour communication. In the limit case of maximal

[4]For instance, take a normal distribution truncated over Ω of mean a and of arbitrary small variance for the optimistic case and a normal distribution truncated over Ω of mean b and of arbitrary small variance for the pessimistic one.

pessimism, m^+ induces action $t(b)$. For ω_0 to be a cut-off type, it must at least be that:

$$\begin{aligned} t(b) &< t^W(\omega_0) \\ \iff \omega_0 &< \frac{b}{N} \end{aligned}$$

Thus, a sufficient condition for no information transmission to be possible is $N > \frac{b}{a}$. In other words, a prior leading to a more prudent total emission level than the uniform one favours the possibility of information transmission between parties. Yet, as before, when societies are large enough, no information transmission is possible.

In the case of small societies, consider the case $\beta = 2$. For ω_0 to be a cut-off type it must be that:

$$\begin{aligned} & t^W(\omega_0) = \frac{t(\omega_0) + t(b)}{2} \\ \iff \quad & \frac{1}{N\omega_0} = \frac{1}{2}(\frac{1}{\omega_0} + \frac{1}{b}) \\ \iff \quad & \omega_0 = b(\frac{2}{N} - 1) \end{aligned}$$

This time the condition $\omega_0 \in (a, b)$ is met if and only if $N < \frac{2b}{a+b}$. Thus, when $N \geq 2$, no information transmission is possible, whatever the support of the receiver's prior.

Concave benefits for emissions— Although I have allowed for varying degrees of convexity in the damage function of the receivers, I have retained the assumption that the benefits of emissions are linear. This is because what matters for the result is the degree of convexity of the costs relative to the benefits of emissions. To see this, consider the more general utility function of the receiver:

$$u_i(e_i, e_{-i}, \omega) = \delta_1 e_i^{\frac{1}{\delta_1}} - \frac{\omega}{\delta_2} \sum_{i=1}^{N} e_i^{\delta_2}$$

where $\delta_1 \geq 1$ and $\frac{1}{\delta_1} < \delta_2$. One can check that u_i is always a concave in e_i.

Then, given any message $m \in \mathcal{M}$, the decentralised level of emissions is:

$$t_i(m) = \frac{N}{\mathbb{E}(\omega|m)^{\frac{\delta_1}{\delta_1\delta_2-1}}}$$

Thus, one can equate β with $\delta_2 - \frac{1}{\delta_1} + 1$ to obtain the same optimal action as before and verify, given the assumptions made on δ_1 and δ_2, that $\beta > 1$. As before, we have that for any message $m \in \mathcal{M}$:

$$t^W(m) = \frac{1}{N^{\frac{1}{\beta-1}}} t(m)$$

As a result, the computation of the potential cut-offs will be the same as before and will therefore not affect the main results. More generally, notice that it would be enough for the benefit and damage functions to be assumed increasing continuous and respectively concave and strictly convex for a general version of the main theorems to hold. Yet, by considering a parametric version of the model, we gained insight on the role of the concavity of these functions and the one played by the number of receivers.

Heterogeneous receivers— I have assumed that all receivers have the same utility function. A natural question is whether my results extend to the case where receivers have heterogeneous preferences. Consider the alternative utility function of receiver i:

$$u_i(e_i, e_{-i}, \omega) = \gamma_i e_i - \frac{\omega}{\beta}\sum_{i=1}^{N} e_i^{\beta}$$

where $\gamma_i > 0$. Here, the socially optimal emission level for R_i, given any message $m \in \mathcal{M}$, becomes:

$$e_i^W(m) = \frac{\gamma_i}{(N\mathbb{E}(\omega|m))^{\frac{1}{\beta-1}}}$$

Thus, in general, the aims of the sender concern not only the total emission level as before, but each receiver's individual emission level as well. In other words, the action variable is now a vector of length N. Formally proving that my results hold true when receivers have heterogeneous preferences is beyond the scope of this paper. However, it has been shown by Goltsman and Pavlov (2008) (Proposition 2) that, in the case of linear biases, communication with two receivers with biases b_1, b_2 is equivalent to communication with a single representative receiver of bias $\frac{b_1+b_2}{2}$. Utility functions with linear biases can be seen as reasonable approximations of the ones used here. One may conjecture that as in Goltsman and Pavlov (2008), in the context of this paper, the communication stage could be equivalent to a one-sender one-receiver game played between a representative receiver whose optimal total emission level under message $m \in \mathcal{M}$ is $t(m) = N\sum_{i=1}^{N}\alpha_i e_i(m)$, where $\sum_{i=1}^{N}\alpha_i = 1$ and $\alpha_i \in [0,1]$ for all $i \in 1, ..., N$, and a representative sender whose optimal total emission level in state $\omega \in \Omega$ is $t^W(\omega) = N\sum_{i=1}^{N}\alpha_i e_i^W(\omega)$. If this is the case, it is easy to see that this paper's main results would be maintained.

References

Angeletos, G.-M. and Pavan, A. (2007). Efficient use of information and social value of information. *Econometrica*, 75(4):1103–1142.

Asheim, G. B. (2010). Strategic use of environmental information. *Environmental and Resource Economics*, 46(2):207–216.

Battaglini, M. (2002). Multiple referrals and multidimensional cheap talk. *Econometrica*, 70(4):1379–1401.

Battaglini, M. et al. (2004). Policy advice with imperfectly informed experts. *Advances in theoretical Economics*, 4(1):1–32.

Bergstrom, T., Blume, L., and Varian, H. (1986). On the private provision of public goods. *Journal of public economics*, 29(1):25–49.

Crawford, V. P. and Sobel, J. (1982). Strategic information transmission. *Econometrica*, pages 1431–1451.

Douenne, T. and Fabre, A. (2020). French attitudes on climate change, carbon taxation and other climate policies. *Ecological Economics*, 169:106496.

Ehret, P. J., Van Boven, L., and Sherman, D. K. (2018). Partisan barriers to bipartisanship: Understanding climate policy polarization. *Social Psychological and Personality Science*, 9(3):308–318.

Everett, J. A., Colombatto, C., Awad, E., Boggio, P., Bos, B., Brady, W. J., Chawla, M., Chituc, V., Chung, D., Drupp, M. A., et al. (2021). Moral dilemmas and trust in leaders during a global health crisis. *Nature human behaviour*, 5(8):1074–1088.

Fischer, A., Peters, V., Vávra, J., Neebe, M., and Megyesi, B. (2011). Energy use, climate change and folk psychology: Does sustainability have a chance? results from a qualitative study in five european countries. *Global Environmental Change*, 21(3):1025–1034.

Gabel, M., Gooblar, J., Roe, C. M., and Morris, J. C. (2021). The ideological divide in confidence in science and participation in medical research. *Scientific reports*, 11(1):1–9.

Galeotti, A., Ghiglino, C., and Squintani, F. (2013). Strategic information transmission networks. *Journal of Economic Theory*, 148(5):1751–1769.

Goltsman, M. and Pavlov, G. (2008). How to talk to multiple audiences.

Halady, I. R. and Rao, P. H. (2010). Does awareness to climate change lead to behavioral change? *International Journal of Climate Change Strategies and Management*.

Hornsey, M. J., Lobera, J., and Díaz-Catalán, C. (2020). Vaccine hesitancy is strongly associated with distrust of conventional medicine, and only weakly associated with trust in alternative medicine. *Social Science & Medicine*, 255:113019.

Kakeu, J. and Johnson, E. P. (2018). Information exchange and transnational environmental problems. *Environmental and Resource Economics*, 71(2):583–604.

Krishna, V. and Morgan, J. (2001). A model of expertise. *The Quarterly Journal of Economics*, 116(2):747–775.

Lorenzoni, I. and Pidgeon, N. F. (2006). Public views on climate change: European and usa perspectives. *Climatic change*, 77(1):73–95.

Maestre-Andrés, S., Drews, S., and van den Bergh, J. (2019). Perceived fairness and public acceptability of carbon pricing: a review of the literature. *Climate Policy*, 19(9):1186–1204.

Meinshausen, M., Meinshausen, N., Hare, W., Raper, S. C., Frieler, K., Knutti, R., Frame, D. J., and Allen, M. R. (2009). Greenhouse-gas emission targets for limiting global warming to 2 c. *Nature*, 458(7242):1158–1162.

Pidgeon, N. and Fischhoff, B. (2011). The role of social and decision sciences in communicating uncertain climate risks. *Nature climate change*, 1(1):35–41.

Slechten, A. (2020). Environmental agreements under asymmetric information. *Journal of the Association of Environmental and Resource Economists*, 7(3):455–481.

Sterman, J. D. (2008). Risk communication on climate: mental models and mass balance. *Science*, 322(5901):532–533.

Tol, R. S. (2011). Regulating knowledge monopolies: the case of the ipcc. *Climatic Change*, 108(4):827.

Van Boven, L., Ehret, P. J., and Sherman, D. K. (2018). Psychological barriers to bipartisan public support for climate policy. *Perspectives on Psychological Science*, 13(4):492–507.

Whitmarsh, L. and Capstick, S. (2018). Perceptions of climate change. In *Psychology and climate change*, pages 13–33. Elsevier.