



**UNIVERSITÀ  
DEGLI STUDI  
DI BERGAMO**

**Department  
of Economics**

## **WORKING PAPERS**

### **What Drives Inaction on Climate Change? A Review of the Literature**

**Francesco Fallucchi, Danièle Fares, Elena Manzoni**  
December 2025 - WP N. 35 Year 2025



Working papers – Department of Economics  
n. 35

# What Drives Inaction on Climate Change? A Review of the Literature



UNIVERSITÀ  
DEGLI STUDI  
DI BERGAMO

Department  
of Economics

Francesco Fallucchi, Danièle Fares,  
Elena Manzoni



---

Università degli Studi di Bergamo  
2025

What Drives Inaction on Climate Change? A Review of the Literature / Francesco Falucchi, Daniele Fares, Elena Manzoni.  
Bergamo: Università degli Studi di Bergamo, 2025.  
Working papers of Department of Economics, n. 35  
ISSN: 2974-5586  
DOI: [10.13122/WPEconomics\\_35](https://doi.org/10.13122/WPEconomics_35)

Il working paper è realizzato e rilasciato con licenza

Attribution Non commercial Non derivatives license (CC BY-NC-ND 4.0)

<https://creativecommons.org/licenses/by-nc-nd/4.0/>

La licenza prevede la possibilità di ridistribuire liberamente l'opera, a patto che venga citato il nome degli autori e che la distribuzione dei lavori derivati non abbia scopi commerciali.



Progetto grafico: Servizi Editoriali – Università degli Studi di Bergamo  
Università degli Studi di Bergamo  
via Salvecchio, 19  
24129 Bergamo  
Cod. Fiscale 80004350163  
P. IVA 01612800167

<https://aisberg.unibg.it/handle/10446/315646>

# What Drives Inaction on Climate Change? A Review of the Literature

Francesco Fallucchi, Danièle Fares, Elena Manzoni

October 2025

## Abstract

Despite widespread concern about climate change, behavioral engagement and policy support remain limited. We present and reinterpret existing evidence through a collective-action framework informed by belief-dependent preferences. Two belief channels—first-order beliefs about others' behavior (descriptive norms) and second-order beliefs about others' expectations (social expectations)—are embedded in a behavioral public-goods model. When beliefs are accurate, these channels sustain conditional cooperation and self-fulfilling collective action. Inaction may instead arise when the belief references are biased downward. We distinguish between two empirically grounded sources of distortion: genuine misperceptions, arising from informational limits and bounded rationality, and motivated misperceptions, driven by self-serving and identity-protective reasoning. This distinction guides policy: visibility and feedback correct genuine errors; identity-compatible framing, in-group messages, and narrative persuasion counter motivated bias. We thus connect the behavioral theory of conditional cooperation with empirical evidence on belief distortions and map the different mechanisms to interventions that overcome collective climate inaction.

**JEL Codes:** *D83, D91, H41, Q54*

**Keywords:** *Environmental behavior, conditional cooperation, beliefs, bounded rationality, motivated reasoning, behavioral interventions*

# 1 Introduction

Despite widespread concern about climate change (Lee et al., 2025; Dechezleprêtre et al., 2025), behavioral engagement and policy implementation remain inadequate. Mitigation is costly, its benefits are diffuse, and success depends on the actions of many others. These features create a persistent intention–action gap: people say they care, yet they often fail to act or to support ambitious policy (Sheeran and Webb, 2016). We study this gap through the lens of collective action with belief-dependent motives.

Our starting point is that climate mitigation constitutes a large-scale collective action problem. The benefits of cooperation are shared, while the costs are borne individually, generating incentives to free-ride. In such settings, many individuals behave as conditional cooperators: they contribute more when they believe others contribute or expect them to do so (Fischbacher et al., 2001). In this respect, two types of beliefs matter. First-order beliefs (FOB)—beliefs about own or others’ actions—play a role by capturing, for example, perceived descriptive norms—what others do. Second-order beliefs (SOB)—beliefs about own or others’ beliefs—matter because they capture perceived expectations—what others think one will or ought to do (guilt aversion motives/injunctive norms). Existing theories often isolate one channel at a time and, crucially, assume that beliefs are accurate. These frameworks explain cooperation under favorable conditions but struggle to account for persistent inaction when concern is high and pro-social intentions are widespread yet invisible.

We propose a behavioral public good model through which we highlight the functioning of each channel, separately but coherently. We assume that individuals care about both their own private consumption and others’ actions or beliefs. Specifically, we assume that deviations below either a descriptive norm based on others’ behavior—FOB channel—or others’ expectations—SOB channel—carry a psychological cost. The resulting best response features an interior matching region in which contributions move one-for-one with the target induced by the relevant belief. When beliefs are accurate, these two mechanisms yield a self-fulfilling cooperative equilibrium: people act because they expect others to act—or expect others to expect them to act—and those expectations are confirmed.

Why, then, does inaction persist? Cooperation collapses when beliefs are systematically biased downward—toward lower norms or expectations. Using belief-dependent models of cooperation as a conceptual framework, we classify the existing literature into two empirically grounded sources of distortion. Genuine misperceptions arise from informational frictions and cognitive limits: mitigation efforts are hard to observe, social feedback is noisy, and people under-infer others’ willingness to act (e.g., pluralistic ignorance). Motivated misperceptions arise from identity-protective or self-serving reasoning: individuals avoid or discount dissonant social information, adopt justificatory narratives, and downweight expectations that would obligate action. After distinguishing between

the cognitive and motivational roots of misperceptions, we examine how they affect the success of interventions aimed at restoring cooperation.

The paper proceeds as follows. Section 2 reviews the theoretical foundations linking beliefs about others to conditional cooperation and connects these mechanisms to empirical evidence from climate-related contexts. Section 3 surveys the relevant literature, describing how misperceptions about others—whether informational or motivated—can shift behavior toward a low-cooperation equilibrium. Section 4 concludes by mapping mechanism to policy: when misperceptions are genuine, visibility and dynamic-norm information are most effective; when they are motivated, identity-compatible framing and in-group messengers are pivotal.

## 2 Conditional Cooperation and Beliefs

Consider a public good game with  $N$  players. Each player  $i$  has an endowment  $w > 0$ , chooses his/her level of contribution to a public good  $g_i$ , and uses the remaining income ( $w - g_i$ ) for private consumption. Let  $G = \sum_{j=1}^N g_j$  denote the total public good, and let  $r \in (0, 1)$  be the marginal per capita return. We assume that the private utility from consumption  $v(w - g_i)$  is increasing and concave ( $v' > 0$ ,  $v'' < 0$ ). For the purpose of the two examples discussed below, we assume  $v(w - g_i) = \ln(w - g_i)$ , which is a convenient specification that ensures analytical tractability.

Player  $i$ 's payoff depends on: i) her private consumption, ii) the level of the public good, and iii) her social preferences on her own contribution level, which may be, for example, preferences for matching others' contributions, or preferences for matching others' expectations. This last component captures *conditional cooperation*: the idea that individuals are more willing to contribute when they believe that others will also contribute.

Let  $\theta_i \geq 0$  denote a parameter capturing how much player  $i$  cares about aligning her contribution with what she thinks is the socially accepted contribution target. Player  $i$ 's utility can be expressed in general form as:

$$u_i(g_i; B_i) = \ln(w - g_i) + rG - \theta_i \Phi(g_i, B_i),$$

where  $\Phi(g_i, B_i)$  measures the disutility from deviating from the socially accepted target, which depends both on player  $i$ 's contribution,  $g_i$ , and on her beliefs  $B_i$ . Note that players make their contributions simultaneously, hence player  $i$  can only rely on her ex-ante beliefs when choosing  $g_i$ .

Different specifications of  $\Phi(\cdot)$ , which may be based on different types of beliefs  $B_i$ , capture alternative social concerns. In what follows, we focus on two examples: (i) compliance with *descriptive social norms*, which depends on first-order beliefs about others' actions (i.e., contributions) (Fehr and Schurtenberger, 2018), and (ii) *guilt aversion*, which depends on

second-order beliefs about others' expectations (Battigalli and Dufwenberg, 2007).

These are illustrative rather than exhaustive examples, and they demonstrate that conditional cooperation may be driven by both first-order beliefs and second-order beliefs. However, these are not the unique channels through which first- and second-order beliefs may generate belief-dependent cooperation. Other well-studied mechanisms include: *Inequity aversion* with disutility from unequal payoffs (Fehr and Schmidt, 1999; Bolton and Ockenfels, 2000); *Reciprocity and fairness intentions* with utility gains from rewarding kind actions or intentions (Rabin, 1993; Dufwenberg and Kirchsteiger, 2004); *Reputation and social image* with concern for how one's actions are perceived by others (Andreoni and Bernheim, 2009; Attanasi et al., 2019). These formulations all share the same underlying logic: individuals derive utility not only from material outcomes but also from how their behavior aligns with beliefs about others' behavior or beliefs.

## 2.1 Two belief-based mechanisms

Before the two illustrative models, let us clarify the distinction between the two types of beliefs. **First-order beliefs** (FOBs) concern what individuals think others do, or what they think about the exogenous parameters of the model; we denote individual  $i$ 's FOBs as  $\alpha_i$ . **Second-order beliefs** (SOBs) concern what agents believe about other agents' beliefs on behavior or exogenous parameters of the model; we denote  $i$ 's SOBs as  $\beta_i$ . Therefore, FOBs are the beliefs that we use when modeling descriptive norms, while SOBs are the relevant beliefs if we want to describe how agents who care about social expectations behave.

We now present the two models, highlighting the link between beliefs and conditional cooperation. All formal derivations, best-responses expressions, and equilibrium conditions are included in Appendix A.

**First-order beliefs and descriptive norms.** Following Fehr and Schurtenberger (2018) and Fallucchi et al. (2022), we consider individuals who experience a psychological cost when their contribution falls short of a prevailing *descriptive norm* of cooperation. The underlying idea is that individuals care about doing “what others typically do”, and deviations below the normative standard are perceived as socially inappropriate. In their model, each player incurs a loss in utility proportional to the squared distance between her contribution  $g_i$  and the norm  $c^*$  whenever she contributes less than  $c^*$ .

If, in the context of the public good game introduced above, we model the norm  $c^*$  as the *descriptive norm of contribution*—the average contribution individual  $i$  expects others to make—the individual's utility depends on FOBs on others' pattern of contribution. Adapting the structure of Fehr and Schurtenberger (2018) to this setting, player  $i$ 's utility becomes:

$$u_i = \begin{cases} \ln(w - g_i) + rG - \theta_i^N (g_i - \frac{1}{N-1} \sum_{j \neq i} g_j)^2, & \text{if } g_i < \frac{1}{N-1} \sum_{j \neq i} g_j, \\ \ln(w - g_i) + rG, & \text{if } g_i \geq \frac{1}{N-1} \sum_{j \neq i} g_j, \end{cases}$$

where  $\theta_i^N \geq 0$  measures the individual's sensitivity to deviations from the norm.

If the individuals want to match others' *current* contribution, and the choices are simultaneous, they cannot observe  $g_j$ , and therefore they have to base their contribution choices on their belief about the *contemporaneous* descriptive norm. To express this formally, let  $\alpha_i^j = \mathbb{E}_i[g_j]$ . The expected norm is therefore  $\bar{\alpha}_i = \frac{1}{N-1} \sum_{j \neq i} \mathbb{E}_i[g_j] = \frac{1}{N-1} \sum_{j \neq i} \alpha_i^j$ , that is, the average of the beliefs that  $i$  has on other  $j$ 's contributions. Hence, individual  $i$ 's utility becomes

$$u_i(g_i, \alpha_i) = \begin{cases} \ln(w - g_i) + rG - \theta_i^N (g_i - \bar{\alpha}_i)^2, & \text{if } g_i < \bar{\alpha}_i, \\ \ln(w - g_i) + rG, & \text{if } g_i \geq \bar{\alpha}_i. \end{cases}$$

The term  $\theta_i^N (g_i - \bar{\alpha}_i)^2$ , therefore, captures the psychological cost of contributing less than the expected average of others. The higher  $\theta_i^N$ , the stronger the discomfort from under-contribution, and thus the stronger the incentive to align one's behavior with the expected norm. This specification interprets conditional cooperation as driven by norm compliance: individuals choose contributions that balance marginal return from the public good with the disutility from norm deviation. When  $\theta_i^N = 0$ , players behave purely selfishly, contributing only when convenient in terms of material returns. As  $\theta_i^N$  increases, they become increasingly more reactive to the perceived descriptive norm; they raise  $g_i$  whenever they expect others to contribute more. In the limit, as  $\theta_i^N \rightarrow \infty$ , the best response mapping approaches the 45-degree line  $g_i = \bar{\alpha}_i$ : players match others' behavior one-to-one.

At the aggregate level, when each individual's belief about others' contributions is correct, i.e., it coincides with actual average behavior, the population will reach a self-consistent cooperative equilibrium, where individuals contribute what they expect others to contribute, and these expectations are confirmed in equilibrium. Conditional cooperation thus emerges endogenously from conformity to the contemporaneous descriptive norm.

**Second-order beliefs and guilt aversion.** Under guilt aversion, individuals care about living up to others' expectations. In contrast to descriptive norms, which rely on FOBs about others' actions, guilt is driven by SOBs: beliefs about others' expectations of one's own contribution (Battigalli and Dufwenberg, 2007). Failing to meet these expectations generates a negative psychological experience—guilt—which individuals seek to avoid. To formalize this idea, recall that  $\alpha_j^i = \mathbb{E}_j[g_i]$  denotes player  $j$ 's expectation about  $i$ 's contribution, and let  $\bar{\alpha}^i = \frac{1}{N-1} \sum_{j \neq i} \alpha_j^i$  represent the aggregate expectation of

player  $i$ 's contribution.

The corresponding utility is

$$u_i(g_i, \bar{\alpha}_i) = \ln(w - g_i) + r \left( g_i + \sum_{j \neq i} g_j \right) - \theta_i^G (\bar{\alpha}_i - g_i)^+$$

where  $\theta_i^G > 0$  captures the degree of guilt sensitivity, and  $(\bar{\alpha}_i - g_i)^+ \equiv \max\{0, \bar{\alpha}_i - g_i\}$ .<sup>1</sup> An individual who expects that others hold high expectations about her contribution will increase  $g_i$  to avoid the disutility from letting them down.

However, because player  $i$  does not observe  $\alpha_{j,i}$ , she forms beliefs about these expectations: let  $\beta_{i,j} = \mathbb{E}_i[\alpha_{j,i}]$  denote player  $i$ 's belief about what player  $j$  expects from her in the public good. The relevant SOB is thus the average expectations she believes others hold about her:  $\bar{\beta}_i = \mathbb{E}_i[\bar{\alpha}_i] = \frac{1}{N-1} \sum_{j \neq i} \beta_{i,j}$ .

The corresponding utility becomes

$$u_i(g_i, \bar{\beta}_i) = \ln(w - g_i) + r \left( g_i + \sum_{j \neq i} g_j \right) - \theta_i^G [\bar{\beta}_i - g_i]^+,$$

where  $[\bar{\beta}_i - g_i]^+ = \max\{0, \bar{\beta}_i - g_i\}$ . This term  $[\bar{\beta}_i - g_i]^+$  is  $i$ 's expectation of the disappointment that he/she generates by contributing less than what others expect. In this formulation, conditional cooperation reflects a desire to meet these perceived expectations. When others do not expect cooperation, i.e.,  $\bar{\beta}_i$  is low, the guilt term is inactive, and behavior aligns with the standard material optimum. As  $\bar{\beta}_i$  increases, the marginal cost of under-contributing increases, giving individuals an incentive to match these expectations.

This formulation interprets conditional cooperation as compliance with the social expectation: individuals contribute not merely because others do, but because they believe others want them to do so. When  $\theta_i^G = 0$ , players behave as standard selfish agents. As  $\theta_i^G$  increases, they become more responsive to others' expectations, raising their contributions when they believe others expect more from them. Within an interior region of beliefs, the best response is thus to match the perceived expectations exactly. When each individual believes that others' expectations align with their own behavior, this process generates the characteristic one-for-one conditional cooperation pattern. Conditional cooperation

---

<sup>1</sup>Note that, while in the descriptive norm specification, the cost was quadratic in  $(g_i - \bar{\alpha}_i)$ , following [Fehr and Schurtenberger \(2018\)](#), here the psychological cost enters linearly through a shortfall function. This difference reflects conventions in the respective literatures rather than substantive modeling differences; the goal here is to illustrate how different belief-based motives generate conditional cooperation, not to compare their functional forms.

thus emerges endogenously from individuals' desire to meet others' expectations.

## 2.2 Beliefs and Conditional Cooperation in Climate Mitigation

Beliefs about others' actions and expectations play a central role in sustaining cooperation in public goods settings (Battigalli and Dufwenberg, 2022). This is particularly true in climate change mitigation, where outcomes depend on collective effort, actions are costly, and feedback is slow, individuals rely on social cues to infer what is expected of them. Observing others' behaviors or perceived expectations signals shared commitment and can reinforce personal engagement (Abrahamse et al., 2005; Allcott, 2011; Farrow et al., 2017). When these beliefs are accurate, conditional cooperation emerges: people act when they believe others are acting or expect them to act, and this sustains a self-fulfilling equilibrium of collective action (Fischbacher et al., 2001; Fischbacher and Gächter, 2010).

Recall that we introduced two types of beliefs: first- and second-order beliefs. Experimental and field evidence confirm that both types of beliefs matter for cooperation. Let us start with the evidence on the effects of FOBs. In laboratory public good games, most participants are *conditional cooperators*: they contribute more when they believe others do so (Fischbacher et al., 2001). This pattern extends to real-world settings: forest communities with more conditional cooperators manage resources more effectively (Rustagi et al., 2010), and fishermen who believe peers fish sustainably are themselves more cooperative (Fehr and Leibbrandt, 2011). Climate-specific experiments reveal the same dynamic: individuals are significantly more likely to purchase CO<sub>2</sub> permits when they believe a majority of others contribute (Sturm et al., 2019). Globally, Andre et al. (2024a) show that perceived willingness to contribute (WTC) strongly predicts actual WTC; each one-point increase in perceived others' WTC raises individual contribution likelihood by nearly 0.5 points.

We now move to discuss the evidence of the effects of SOBs, which are especially influential when cooperation is costly or unobservable. Experiments in energy conservation show that people reduce consumption more when they believe others support conservation than when they merely believe in climate facts (Jachimowicz et al., 2018). Likewise, individuals in the U.S. and China are more likely to back climate policies if they believe peers or in-group members support them, regardless of personal concern (Mildenberger and Tingley, 2019). These findings suggest that perceived expectations, more than private conviction, motivate sustained pro-environmental behavior.

Beliefs about others' expectations are shaped through both *top-down* and *bottom-up* channels. Regarding the top-down channel, political and media cues often convey what is socially expected within one's group: people interpret elite endorsements as signals of their in-group's expectations (Ehret et al., 2018; Abeles et al., 2019). Yet, the most powerful

beliefs arise from close social networks. Conversations with trusted peers and family strengthen perceptions of consensus and shared responsibility (Goldberg et al., 2019). Schuldt et al. (2019) show that SOBs about “friends and family” predict climate policy support far more strongly than beliefs about the general population, while Goldberg et al. (2020) find that perceiving support among one’s social circle narrows partisan divides on climate policy. These results align with social identity theory (Tajfel et al., 1979; Hornsey, 2008): individuals internalize expectations more readily when they come from similar or trusted others.

Hence, we can conclude that climate mitigation thus depends on the accuracy of beliefs about others’ cooperation and expectations. When these beliefs are correct, the equilibrium of conditional cooperation sustains itself. When they are distorted—either by lack of information or by motivated reasoning—the system drifts toward a low-cooperation equilibrium.

### 3 From correct to distorted beliefs: when conditional cooperation fails

The previous framework shows that conditional cooperation arises when individuals hold accurate beliefs about others’ contributions or expectations. With correct beliefs, equilibrium cooperation is self-fulfilling: people act because they expect others to act, and these expectations are confirmed. In reality, however, belief-action consistency rarely holds. Information about others is incomplete or selectively interpreted, producing systematic departures from cooperative equilibrium. We distinguish between two origins of belief distortion: **genuine misperceptions**, rooted in informational deficit or cognitive constraints, and **motivated misperceptions**, shaped by self-serving or identity-protective reasoning. While both mechanisms distort beliefs about others, they do so via distinct mechanisms and have different implications for policy interventions, hence it is important to distinguish when either type of misperception is in place.

#### 3.1 Genuine Misperceptions

A misperception, in this context, refers to a systematic deviation between an individual’s belief about others’ behaviors or expectations of them and the actual distribution of those behaviors or expectations within the relevant social environment. This section examines genuine misperceptions, where distortions are best understood as arising from bounded rationality or limited information: individuals operate within constraints of limited information, cognitive capacity, and inferential sophistication (Simon, 1955; Tversky and Kahneman, 1974). Because climate mitigation efforts are hard to observe and collective

outcomes unfold slowly, people often underestimate others’ willingness to cooperate. This pessimism discourages their own participation and generates a self-confirming low-effort equilibrium: everyone expects little, contributes little, and thereby validates the expectations.

### 3.1.1 Bounded Rationality and the Formation of Genuine Misperceptions

**Information and Observability Gaps** In the early stages of learning or when information is limited, individuals act on possibly incorrect beliefs: they choose best responses to their own subjective expectations rather than to actual contributions. In climate domains, individual actions such as dietary change, household energy efficiency, or policy support are largely invisible. As [Bicchieri \(2005\)](#) argues, the absence of observability weakens individuals’ ability to infer social expectations because expectations cannot be anchored to visible behavior. Without clear signals that others are acting or expecting action, individuals may infer that no such expectations exist. Exposure to homogeneous peer or media environments further narrows the informational field, producing echo chambers that only amplify local norms ([Jasny et al., 2015](#)). Without credible feedback, individuals fill the gap with pessimistic priors about collective engagement, creating a situation of “informational deficit” regarding others’ cooperative intentions ([Suldozsky, 2017](#)).

**Cognitive Constraints** Even when information is available, forming accurate beliefs is cognitively demanding because it requires recursive reasoning about how others interpret information and about how their behavior signals these interpretations. Such recursive reasoning is often modeled as level- $k$  reasoning ([Camerer et al., 2004](#); [Costa-Gomes and Crawford, 2006](#)). Evidence from climate-relevant public goods experiments shows that individuals often struggle with such higher-order inference. In a study using threshold climate games, participants systematically failed to anticipate how others condition their contributions on shared risks, leading to coordination failures that resemble low levels of recursive reasoning ([Tavoni et al., 2011](#)). Similarly, [Guarino and Jehiel \(2013\)](#) show that individuals engaged in social learning often rely on coarse inference: rather than forming beliefs over the full strategic structure of the environment, they group others’ behaviors into a limited set of coarse categories, which leads them to overweight salient signals and systematically misinterpret what others’ actions imply about their underlying intentions or expectations. These limitations are amplified in the climate domain, where individuals frequently default to simplifying heuristics, selective attention, or confirmation-driven processing when interpreting social information about climate attitudes and behaviors ([Frank et al., 2024](#)). Moreover, the ability to integrate multiple perspectives—a marker of cognitive complexity—varies across individuals, and lower complexity is associated with

more shallow processing of climate-related arguments (Chen and Unsworth, 2019). As a result, many people form noisy or biased beliefs about others, not because information is unavailable, but because accurately inferring what others believe or expect exceeds their cognitive capacity or motivation.

**Heuristic-based errors: Pluralistic ignorance and False Consensus** In complex or opaque environments, people use shortcuts that systematically distort social inference. Two particularly relevant mechanisms in the climate context are pluralistic ignorance and the false consensus effect. Pluralistic ignorance arises when a majority of members of a group or society privately have a certain belief, opinion, or practice, yet believe that practically every other member thinks the opposite (Katz et al., 1931; Miller and McFarland, 1987; Prentice and Miller, 1996).<sup>2</sup> Individuals may privately endorse climate action yet believe others do not. Recent evidence shows that U.S. citizens significantly underestimate national support for key climate policies, perceiving only 37–43% support when actual levels range from 68–80% (Sparkman et al., 2022). This misperception is especially pronounced in partisan contexts: Republicans underestimate climate concern among fellow Republicans, largely due to perceived identity mismatches (Lyons and Hasell, 2024). Similar findings emerge in occupational settings: Reynaud and Ouvrard (2024) document that French farmers systematically underestimate their peers’ support for eco-schemes under the Common Agricultural Policy, reducing willingness to adopt them. Specifically, farmers underestimated by 15.3% to 15.5% the proportion of peers who believed in the environmental benefits of these schemes, and by 27.8% to 30.9% the proportion willing to adopt them. At a global level, Andre et al. (2024a,b) document a pattern of underestimating both others’ willingness to act and others’ beliefs about what should be done to mitigate climate change, despite private support. When experimental participants were shown actual distributions of beliefs and willingness to contribute, both monetary donations and policy support increased, particularly among skeptics.

Conversely, the false consensus effect leads individuals to overestimate the degree to which their own beliefs are shared (Ross et al., 1977; Marks and Miller, 1987). Climate skeptics are particularly prone to this bias (Sanders and Mullen, 1983), especially in ideologically fragmented or polarized environments. Mildenberger and Tingley (2019), using nationally representative samples from the U.S. and China, find that individuals consistently underestimate the proportion of pro-climate attitudes among their fellow citizens, and that climate skeptics in particular overestimate public agreement with their views. Importantly, this misperception is observed not only among the general public but also among the political and business elite, underscoring its reach across institutional levels. Similarly, Leviston et al. (2013), in a longitudinal panel survey in Australia, show

---

<sup>2</sup>See Bursztyn et al. (2020) and Bursztyn and Yang (2022) for examples of pluralistic ignorance in other contexts.

that climate skeptics vastly overestimate the presence of skepticism, believing that nearly half the population shares their views, when in fact only 7.2% do so. At the same time, pro-climate individuals tend to underestimate public support, highlighting how false consensus and pluralistic ignorance can coexist within the same belief environment. This coexistence is also shown by [Sokoloski et al. \(2018\)](#), in a study about public attitudes toward offshore wind infrastructure. In one sample, supporters underestimated support for renewable development, while opponents overestimated it. This resulting information asymmetry distorted public debate and gave the false impression of low consensus. Indeed, false consensus and pluralistic ignorance reinforce one another. Skeptics, by being more vocal or overrepresented in media discourse, create the illusion of majority skepticism. These biases coexist and reinforce each other: vocal minorities create the illusion of widespread skepticism, while supportive majorities remain silent, eroding the perceived consensus necessary for norm-driven cooperation ([Dixon et al., 2024](#)).

### 3.1.2 From Best Responses to Self-Confirming Equilibria

The mechanisms above explain why individuals enter cooperation dynamics based on inaccurate social beliefs. In practice, people do not solve for a full rational-expectations equilibrium; they follow *best-response thinking*: acting on the basis of their subjective beliefs about others rather than on actual contributions. As long as those beliefs roughly match observed outcomes, behavior appears consistent and persists. With limited or biased feedback, these subjective equilibria become self-sustaining. Under such conditions, agents are rational: each player’s strategy is a best response to some internally consistent belief. In the context of climate cooperation, this implies that if individuals expect little contribution, observe little visible action, and thus find their pessimism confirmed: this is a stable but inefficient low-contribution equilibrium. In the framework introduced previously, players still best respond to belief-based references ( $N_i$  or  $\bar{\beta}_i$ ), but because beliefs are downward-biased, equilibrium cooperation remains trapped at a low level.

Moreover, the presence of sparse and noisy feedback about others’ cooperation may make beliefs converge only locally, producing what [Fudenberg and Levine \(1993\)](#) call a *self-confirming equilibrium*. Social communication dynamics are a potential cause of these informational distortions. When people believe that few others care or act, they hesitate to express their own pro-environmental views, reinforcing the appearance of apathy. This process, known as the *spiral of silence* ([Noelle-Neumann, 1974](#)), creates a visibility bias that sustains pessimistic expectations: widespread private support remains hidden because individuals fear social isolation or reputational costs. The resulting communication gap strengthens the self-confirming loop: beliefs of inaction suppress the expression and visibility of action, which in turn confirms the initial belief. [Maibach et al. \(2016\)](#) report misperceptions hinder people’s willingness to talk about global warming: 57% of Americans rarely or never talk about global warming, and nearly a quarter never hear it dis-

cussed, despite broad concern. This creates a spiral of silence, where the belief that climate inaction is widespread reduces pro-climate opinions, creating false perceptions of social expectations and maintaining the status quo of inaction. [Geiger and Swim \(2016\)](#) experimentally confirmed this mechanism: participants who underestimated pro-climate beliefs in their peer group were less likely to share their own views, and an over-representation of individuals who held anti-climate beliefs, who, on the contrary, believe their opinion is more common than they think, dominated the conversation. The authors find that a spiral of silence develops when pro-environment individuals, who exhibit pluralistic ignorance by underestimating the proportion of the population that holds pro-climate beliefs, are less likely to communicate their beliefs to others.

Overall, genuine misperceptions reflect informational and cognitive limits rather than strategic bias. Yet their behavioral implications mirror those of strategic defection: by weakening the perception of shared effort, they erode the belief-action feedback on which conditional cooperation depends.

## 3.2 Motivated Misperceptions

While genuine misperceptions arise from informational or cognitive constraints, motivated misperceptions reflect directional and self-serving distortions of belief. They emerge when individuals avoid, reinterpret, or strategically construct social information in ways that preserve a convenient view of others' behavior and expectations. In the context of climate change, these distortions enable individuals to justify inaction and free-riding, even when evidence indicates broad social support for mitigation. Motivated misperceptions, therefore, represent a departure from equilibrium not because information is missing, but because individuals are psychologically or socially motivated to resist it. By altering the belief-based reference that sustains conditional cooperation, they erode the reciprocal logic of "I will act if others act or expect me to."

### 3.2.1 Motivated Beliefs and Conditional Cooperation

Motivated beliefs are shaped less by evidence than by psychological or social needs: to avoid cognitive dissonance, maintain self-consistency, or protect group identity. They arise through motivated reasoning: the selective acquisition and interpretation of information to reach preferred conclusions rather than accurate ones ([Kunda, 1990](#); [Epley and Gilovich, 2016](#)). As the "Cultural Cognition Thesis" demonstrates, individuals' perceptions of others' expectations often reflect the norms of their cultural or political in-group more than an unbiased reading of facts ([Kahan et al., 2012](#)). Motivated beliefs thus diverge systematically from reality, yet remain subjectively coherent and socially functional.

These distortions can be self-serving or identity-protective. *Self-serving* motivated beliefs rationalize costly or unpopular behavior while preserving moral self-image. Indi-

viduals downplay others’ cooperation (“most people do nothing”) or reinterpret social expectations as lenient, providing a credible rationale for inaction. In public discourse, such “justifying rationales” help minimize reputational costs while sustaining self-consistency (Bursztyn et al., 2023). *Group-serving* motivated beliefs, by contrast, affirm in-group norms or superiority, sustaining cohesion but distorting reciprocity. For instance, people often believe that their own group contributes more to climate mitigation than others, irrespective of evidence (Taddicken et al., 2019). Both forms undermine conditional cooperation: when perceived expectations ( $\bar{\beta}_i$ ) or descriptive norms ( $\bar{\alpha}_i$ ) are strategically downweighted, equilibrium contributions  $g_i^*$  decline even though the belief–response link remains positive.

### 3.2.2 Mechanisms of Motivated Belief Distortion

**A Priori Avoidance: Strategic Ignorance of Social Expectations** One route to motivated misperception is the deliberate avoidance of social information that could increase perceived obligation to act. Individuals may choose not to learn what others do or expect, maintaining ignorance that shields them from moral or social pressure (Grossman and Van Der Weele, 2017; Golman et al., 2017). In climate domains—where mitigation behaviors are effortful—such *a priori* motivated ignorance preserves a convenient status quo. This avoidance aligns with the logic of cognitive dissonance reduction (Festinger, 1962): rather than reconciling personal behavior with pro-environmental norms, individuals minimize discomfort by remaining uninformed or selectively engaging with confirmatory sources.

Empirical evidence supports this mechanism. Rimbaud and Soldà (2024) find that in trust games, individuals avoid learning about others’ expectations when doing so could induce moral conflict with monetary incentives. Similarly, Dimant et al. (2024) show that people prefer norm information from lenient in-group sources, avoiding stricter signals that could generate moral obligations. In climate communication, selective exposure to ideologically congruent media creates echo chambers that filter out cross-cutting cues (Jasny et al., 2015; Del Vicario et al., 2017). Online discussions reinforce this bias: both climate skeptics and advocates engage primarily with sources aligned with their views, selectively engaging not only with facts but also with others’ perceived beliefs (Areni, 2024). Such environments sustain a form of “motivated Bayesianism” (Gino et al., 2016), where the updating process never begins because dissonant evidence is never encountered.

**Motivated Distortion During Belief Updating** Even when individuals are exposed to social information, motivated reasoning can distort how it is processed. Rather than ignoring cues, individuals reinterpret them defensively, discounting their credibility or reframing them to maintain preferred conclusions (Kunda, 1990). This process, often described as identity-protective cognition, leads people to trust sources aligned with

their in-group and reject others, resulting in biased but internally consistent updating (Druckman and McGrath, 2019; Bago et al., 2023; Bayes and Druckman, 2021). Confirmation and disconfirmation biases reinforce this tendency (Nickerson, 1998; Golman et al., 2016). In the climate domain, where attitudes are highly politicized, individuals process information about others’ expectations through the filter of group identity (Hart and Nisbet, 2012; Doell et al., 2021). Conservatives, for instance, discount messages from environmental institutions perceived as “left-leaning,” while liberals ignore cues framed by conservative messengers (Cohen, 2003; Fielding et al., 2012). Such asymmetric trust produces a “negative cooperation” dynamic: each group sees the other as uncooperative, reinforcing inaction through mutual distrust.

A large body of experimental evidence shows that individuals also strategically distort beliefs to preserve moral self-image under informational ambiguity. When the social consequences of one’s actions are uncertain, people exploit this “moral wiggle room” to act selfishly while maintaining plausible deniability. In dictator and public-goods settings, Dana et al. (2007) demonstrate that reduced transparency lowers giving precisely because participants can deny responsibility. Similarly, Di Tella et al. (2015) find that individuals systematically downplay others’ altruism to justify their own lower contributions, even when accuracy is incentivized. Such distortions occur without conscious deceit: beliefs are adjusted in the direction that renders one’s behavior norm-consistent.

Motivated distortion thus replaces the equilibrium of mutual expectation with one of mutual skepticism. In the framework we developed, the functional form  $g_i = BR_i(B_i)$  still applies, but  $B_i$ —the belief-based reference—is endogenously biased downward by identity-protective filtering. Individuals act conditionally, yet their conditions are defined by pessimistic, self-justifying beliefs. Those misperceptions are sustained not by lack of information but by the psychological and social utility of believing them. The resulting state mirrors a self-confirming equilibrium: each player’s belief about others’ expectations is validated by the very inaction those beliefs sustain.

### 3.2.3 Narratives as Amplifiers of Motivated Beliefs

Narratives provide the interpretive infrastructure that anchors and spreads motivated misperceptions. They transform isolated biases into coherent, emotionally resonant stories that justify inaction while preserving moral coherence (Shiller, 2017; Fløttum and Gjerstad, 2017; Tuckett and Nikolic, 2017; Roos and Reccius, 2024). In climate discourse, these narratives often take the form of “climate delay” or “free-rider” arguments, portraying others as insincere (“hypocrisy” narratives) or emphasizing that mitigation should come from others first (Lamb et al., 2020; Hornsey and Fielding, 2020; Cherry et al., 2024). Such narratives lower perceived collective expectations and weaken reciprocity, functioning as cognitive and social justifications for defection.

The diffusion of these narratives is amplified by polarized media ecosystems and social

media algorithms that prioritize engagement over accuracy (Brady et al., 2017; Falkenberg et al., 2022). Online, emotionally charged and identity-consistent content circulates more rapidly than corrective information, while automated accounts and coordinated campaigns create false signals of public opinion (Ferrara et al., 2016). Empirical studies show that contrarian narratives surged after COP26, reinforcing climate polarization (Falkenberg et al., 2022). In such environments, motivated misperceptions become collectively entrenched: individuals no longer merely misread others’ expectations—they inhabit narrative worlds that make these misperceptions appear true.

### 3.2.4 Motivated Misperceptions as Endogenous Departures from Equilibrium

Motivated misperceptions constitute endogenous departures from belief–action consistency. Unlike genuine misperceptions, which reflect incomplete learning, motivated distortions stem from preferences over which beliefs to hold. They transform the equilibrium condition itself: agents still best respond to perceived norms and expectations, but those perceptions are endogenously biased downward through self-serving cognition and identity alignment. The equilibrium that emerges is internally coherent yet collectively inefficient—a belief-consistent low-cooperation state stabilized by psychological utility rather than informational accuracy. Restoring cooperation, therefore, requires more than correcting misinformation: it demands interventions that make pro-social identities and expectations salient without threatening group belonging.

## 4 Restoring Equilibrium: What Works and What Does Not

Conditional cooperation depends both on belief-action consistency and on the accuracy of beliefs. When individuals hold accurate beliefs about others’ cooperation or expectations, their preferences for conditional cooperation translate into collective action: each person’s contribution reinforces others’ expectations, sustaining a self-fulfilling cooperative equilibrium. In climate change mitigation, where individual effort is costly and outcomes are diffuse, this mechanism is essential. Collective progress depends not only on what people do, but also on what they believe others do and expect.

The preceding sections showed how this equilibrium breaks down. Genuine misperceptions arise from informational and cognitive limits: people lack reliable cues about others’ behavior or expectations and thus underestimate collective engagement. Motivated misperceptions, by contrast, are not errors of ignorance but products of self-serving or identity-protective reasoning. Both undermine conditional cooperation, yet they do so

through distinct channels: the first through incomplete information, the second through motivated interpretation of that information.

Restoring cooperation thus requires different levers. When misperceptions are genuine, the most effective remedies are informational. Increasing visibility and feedback about others' engagement can correct pessimistic beliefs and restore confidence in mutual participation. Experimental evidence confirms that revealing the true distribution of beliefs and willingness to act increases pro-climate engagement, particularly when information is credible and value-congruent (Andre et al., 2024a,b; Goldberg et al., 2019; Schuldt et al., 2019). Highlighting dynamic norms—that more people are beginning to act—can further convey normative momentum and help realign expectations (Sparkman and Walton, 2017; Sabherwal et al., 2021). However, factual data alone often fails to persuade: individuals respond more strongly when social information is embedded in relatable narratives or examples from their community (Bushell et al., 2017; Yang and Hobbs, 2020).

When misperceptions are motivated, informational corrections alone are insufficient. The obstacle is not the absence of knowledge but psychological resistance. Individuals interpret evidence through identity-protective and self-justifying filters, rejecting messages that imply moral obligation or ideological disloyalty (Hart and Nisbet, 2012; Bayes et al., 2020; Ma et al., 2019). Here, effective interventions must work through the psychological channel. Framing climate action in identity-compatible terms, such as energy independence, innovation, or community pride, reduces perceived threat and fosters internal alignment (Feinberg and Willer, 2013; Druckman and McGrath, 2019; Bayes et al., 2020). Trusted in-group messengers and positive norm reframing can make cooperation appear both expected and desirable (Kahan et al., 2012). Similarly, narrative-based persuasion can convey expectations indirectly, through emotionally engaging stories about relatable individuals taking climate action, thereby sustaining belief–action coherence without overt moral pressure (Barron and Fries, 2024; Bushell et al., 2017). Emotional cues such as hope, pride, or shared anger can reinforce perceptions of shared commitment, whereas fear-based or “doom” narratives may exacerbate apathy (Chapman et al., 2017; Hinkel et al., 2020; Moser, 2010).

Across both domains, the goal is not merely to increase knowledge but to rebuild shared beliefs and mutual expectations. Successful collective action requires a belief-based equilibrium in which people's perceptions of others' behavior and expectations are accurate and mutually reinforcing. Restoring cooperation thus entails more than transmitting correct information; it requires rebuilding trust in others' engagement and aligning perceived norms with actual collective intentions.

# A Appendix: Formal Analysis of First- and Second-Order Belief Mechanisms

This appendix provides a detailed derivation of best responses and symmetric equilibria under the two preference specifications used in the main text. Throughout, we assume a one-shot public good game with  $n$  players, each choosing contribution  $g_i \in (0, w)$  from an exogenous income  $w > 0$ . The utility from public good consumption is linear with marginal per-capita return  $r \in (0, 1)$ , and individual utility includes a belief-dependent term capturing either descriptive-norm compliance (first-order beliefs) or guilt aversion (second-order beliefs).

## First-order beliefs and descriptive norms

**Utility and interpretation.** Recall that under the descriptive-norm specification, individual  $i$ 's utility is

$$u_i(g_i, \bar{\alpha}_i) = \ln(w - g_i) + r \left( g_i + \sum_{j \neq i} g_j \right) - \theta_i^N (g_i - \bar{\alpha}_i)^2, \quad (1)$$

**Best response.** To derive the best response, fix  $g_{-i}$  and  $\bar{\alpha}_i$  and differentiate (1) with respect to  $g_i$ :

$$\frac{\partial u_i}{\partial g_i} = -\frac{1}{w - g_i} + r - 2\theta_i^N (g_i - \bar{\alpha}_i). \quad (2)$$

An interior optimum  $g_i \in (0, w)$  satisfies the first-order condition

$$-\frac{1}{w - g_i} + r - 2\theta_i^N (g_i - \bar{\alpha}_i) = 0. \quad (2')$$

The second derivative is

$$\frac{\partial^2 u_i}{\partial g_i^2} = -\frac{1}{(w - g_i)^2} - 2\theta_i^N < 0,$$

so any solution to (2) is indeed a unique local maximum. This defines an implicit best-response function  $BR_i(\bar{\alpha}_i)$  as the unique  $g_i$  solving (2) for given  $\bar{\alpha}_i$ .

Although (2) does not yield a simple closed-form expression for  $BR_i(\bar{\alpha}_i)$ , its comparative statics with respect to the perceived norm are transparent. Define

$$F(g_i, \bar{\alpha}_i) \equiv -\frac{1}{w - g_i} + r - 2\theta_i^N (g_i - \bar{\alpha}_i).$$

Then  $F(g_i, \bar{\alpha}_i) = 0$  implicitly defines  $g_i = BR_i(\bar{\alpha}_i)$ . By the implicit function theorem,

$$\frac{\partial BR_i}{\partial \bar{\alpha}_i} = -\frac{\partial F / \partial \bar{\alpha}_i}{\partial F / \partial g_i} = -\frac{2\theta_i^N}{-\frac{1}{(w - g_i)^2} - 2\theta_i^N}.$$

The denominator is strictly negative, so for  $\theta_i^N > 0$  we obtain

$$\frac{\partial BR_i}{\partial \bar{\alpha}_i} > 0. \quad (3)$$

Thus the best response is strictly increasing in the perceived descriptive norm: as  $i$  believes others contribute more on average, she optimally increases her own contribution.

Two benchmark cases are useful for interpretation:

- If  $\theta_i^N = 0$ , the norm term drops out and (2) reduces to

$$-\frac{1}{w - g_i} + r = 0 \quad \Rightarrow \quad g_i = w - \frac{1}{r},$$

i.e. the standard selfish optimum (provided  $w > 1/r$ ).

- As  $\theta_i^N \rightarrow \infty$ , the quadratic penalty dominates and the optimum approaches  $g_i = \bar{\alpha}_i$ : the agent fully matches the descriptive norm.

For intermediate  $\theta_i^N$ , the best response trades off selfish incentives and norm compliance, and (3) shows that the strength of norm sensitivity determines how responsive contributions are to shifts in the perceived norm.

**Symmetric equilibrium and conditional cooperation.** We now characterize symmetric Nash equilibria under homogeneous parameters  $(w, r, \theta^N)$  and discuss their conditional cooperation property.

**Proposition 1 (Descriptive norms and conditional cooperation)** *Assume that all individuals have utility (1), and that they are homogeneous in  $(w, r, \theta^N)$  with  $\theta^N > 0$ . Let  $BR(\bar{\alpha})$  denote the best-response function defined by (2). Then:*

1.  $BR(\bar{\alpha})$  is strictly increasing in  $\bar{\alpha}$  for all  $\bar{\alpha} \in (0, w)$ .
2. In any symmetric Nash equilibrium, contributions  $g^*$  satisfy

$$g^* = BR(g^*),$$

*and if agents coordinate on a higher perceived descriptive norm  $g' > g^*$ , the corresponding symmetric equilibrium contribution  $g'$  is higher.*

**Proof** Part (1) follows directly from (3), which shows  $\partial BR / \partial \bar{\alpha} > 0$  for  $\theta^N > 0$ .

For part (2), a symmetric Nash equilibrium is a profile  $(g_1, \dots, g_n)$  with  $g_i = g^*$  for all  $i$ , such that each  $g^*$  maximizes  $u_i$  given others' contributions. Under symmetry,  $i$ 's perceived average contribution of others is  $\bar{\alpha} = g^*$ , and the best-response condition reduces to

$$g^* = BR(g^*).$$

Monotonicity of  $BR$  then implies that if players instead coordinate on a higher perceived norm  $\bar{\alpha}' = g' > g^*$  and best-respond to it, the resulting symmetric fixed point  $g' = BR(g')$  must satisfy  $g' > g^*$ . ■

Note that equilibrium contributions are increasing in the norm around which beliefs coordinate. Hence, the mechanism exhibits conditional cooperation: higher perceived average contributions by others induce higher equilibrium contributions by each player.

## Second-order beliefs and guilt aversion

**Utility and interpretation.** We now turn to the guilt-aversion specification, in which the relevant belief is a second-order beliefs about what others expect from  $i$ . Individual utility is

$$u_i(g_i, \bar{\beta}_i) = \ln(w - g_i) + r \left( g_i + \sum_{j \neq i} g_j \right) - \theta_i^G (\bar{\beta}_i - g_i)^+, \quad (4)$$

**Best response.** We first characterize the best response to a given expectation  $\bar{\beta}_i$ . There are two regions, depending on whether the guilt term is active.

*Case 1:  $g_i \geq \bar{\beta}_i$  (inactive penalty).* In this region,  $(\bar{\beta}_i - g_i)^+ = 0$  and the utility reduces to

$$u_i(g_i, \bar{\beta}_i) = \ln(w - g_i) + r \left( g_i + \sum_{j \neq i} g_j \right).$$

The first-order condition for an interior optimum is

$$-\frac{1}{w - g_i} + r = 0 \quad \Rightarrow \quad g_i = g_i^* \equiv w - \frac{1}{r},$$

provided  $g_i^* \geq \bar{\beta}_i$  and  $g_i^* \in (0, w)$ .

*Case 2:  $g_i < \bar{\beta}_i$  (active penalty).* In this region,  $(\bar{\beta}_i - g_i)^+ = \bar{\beta}_i - g_i$  and the utility becomes

$$u_i(g_i, \bar{\beta}_i) = \ln(w - g_i) + r \left( g_i + \sum_{j \neq i} g_j \right) - \theta_i^G (\bar{\beta}_i - g_i).$$

The first-order condition is now

$$-\frac{1}{w - g_i} + (r + \theta_i^G) = 0 \quad \Rightarrow \quad g_i = g_i^{**} \equiv w - \frac{1}{r + \theta_i^G},$$

provided  $g_i^{**} < \bar{\beta}_i$  and  $g_i^{**} \in (0, w)$ . Note that  $g_i^{**} > g_i^*$  because  $r + \theta_i^G > r$ .

These two benchmarks  $g_i^*$  and  $g_i^{**}$  are the selfish optimum (no guilt) and the “guilt-adjusted” optimum (when falling short of expectations is costly), respectively.

Combining the two regions yields the piecewise best response to a given expectation  $\bar{\beta}_i$ :

$$BR_i(\bar{\beta}_i) = \begin{cases} g_i^*, & \text{if } \bar{\beta}_i \leq g_i^*, \\ \bar{\beta}_i, & \text{if } g_i^* < \bar{\beta}_i < g_i^{**}, \\ g_i^{**}, & \text{if } \bar{\beta}_i \geq g_i^{**}. \end{cases} \quad (5)$$

When expectations are low ( $\bar{\beta}_i \leq g_i^*$ ), the guilt term doesn’t bind and the agent behaves selfishly. When expectations are moderate ( $\bar{\beta}_i \in (g_i^*, g_i^{**})$ ), the agent finds it optimal to exactly meet the perceived expectation, so contributions track expectations one-for-one. When expectations are very high ( $\bar{\beta}_i \geq g_i^{**}$ ), the agent is willing to contribute more than the selfish optimum but caps her contribution at  $g_i^{**}$  even if expectations increase further.

**Symmetric equilibrium and conditional cooperation.** Under the homogeneity and symmetry assumptions, a symmetric rational-expectations equilibrium requires that each player’s contribution equals what she expects others to expect from her, so  $g_i = g$  and  $\bar{\beta}_i = g$  for all  $i$ . We can now characterize the set of such equilibria.

**Proposition 2 (Guilt aversion and conditional cooperation)** *Suppose all players share the same  $(w, r, \theta^G)$  and utility (4), and suppose each player expects a symmetric contribution level from everyone else, so that  $\bar{\beta}_i = g$  for all  $i$ . Let  $g^* = w - 1/r$  and  $g^{**} = w - 1/(r + \theta^G)$  with  $0 < g^* < g^{**} < w$ . Then:*

1. *The best-response function  $BR(\bar{\beta})$  given by (5) is weakly increasing in  $\bar{\beta}$ .*
2. *For any  $g \in [g^*, g^{**}]$ , the symmetric profile with  $g_i = g$  and  $\bar{\beta}_i = g$  for all  $i$  is a Nash equilibrium. In particular, for any  $g \in (g^*, g^{**})$ , we have  $BR(g) = g$ , so expected contributions and actual contributions coincide.*

**Proof** Part (1) follows immediately from the piecewise definition (5):  $BR(\bar{\beta})$  is constant on  $(-\infty, g^*]$ , equal to the identity on  $(g^*, g^{**})$ , and constant again on  $[g^{**}, \infty)$ , hence weakly increasing.

For part (2), consider any  $g \in [g^*, g^{**}]$  and suppose that each player  $i$  expects  $\bar{\beta}_i = g$  and chooses  $g_i = g$ . If  $g = g^*$ , then  $\bar{\beta}_i = g^*$  and (5) yields  $BR(\bar{\beta}_i) = g^*$ , so  $g_i = g^*$  is a best response. If  $g = g^{**}$ , then  $\bar{\beta}_i = g^{**}$  and (5) yields  $BR(\bar{\beta}_i) = g^{**}$ , so  $g_i = g^{**}$  is a best response. If  $g \in (g^*, g^{**})$ , then  $\bar{\beta}_i \in (g^*, g^{**})$  and (5) gives  $BR(\bar{\beta}_i) = \bar{\beta}_i = g$ , so again  $g_i = g$  is a best response. In all cases, no player has a profitable unilateral deviation, so the profile is a Nash equilibrium. ■

Note that, within the intermediate region  $g \in [g^*, g^{**}]$ , the guilt mechanism generates conditional cooperation: higher expected contributions  $g$  lead players to choose exactly

higher contributions  $g$ . There, (5) implies  $BR(\bar{\beta}) = \bar{\beta}$ , so if players coordinate on a higher expected level  $g' > g$  within that range, their actual contributions also increase one-for-one from  $g$  to  $g'$ . Thus, in this region, the guilt mechanism exhibits a strong form of conditional cooperation: higher second-order beliefs about what others expect from you translate directly into higher contributions.

## References

- ABELES, A. T., L. C. HOWE, J. A. KROSNICK, AND B. MACINNIS (2019): “Perception of public opinion on global warming and the role of opinion deviance,” *Journal of Environmental Psychology*, 63, 118–129.
- ABRAHAMSE, W., L. STEG, C. VLEK, AND T. ROTHENGATTER (2005): “A review of intervention studies aimed at household energy conservation,” *Journal of Environmental Psychology*, 25, 273–291.
- ALLCOTT, H. (2011): “Social norms and energy conservation,” *Journal of Public Economics*, 95, 1082–1095.
- ANDRE, P., T. BONEVA, F. CHOPRA, AND A. FALK (2024a): “Globally representative evidence on the actual and perceived support for climate action,” *Nature Climate Change*, 14, 253–259.
- (2024b): “Misperceived social norms and willingness to act against climate change,” *Review of Economics and Statistics*, 1–46.
- ANDREONI, J. AND B. D. BERNHEIM (2009): “Social image and the 50–50 norm: A theoretical and experimental analysis of audience effects,” *Econometrica*, 77, 1607–1636.
- ARENI, C. S. (2024): “Motivated reasoning and climate change: Comparing news sources, politicization, intensification, and qualification in denier versus believer subreddit comments,” *Applied Cognitive Psychology*, 38, e4167.
- ATTANASI, G., P. BATTIGALLI, E. MANZONI, AND R. NAGEL (2019): “Belief-dependent preferences and reputation: Experimental analysis of a repeated trust game,” *Journal of Economic Behavior & Organization*, 167, 341–360.
- BAGO, B., D. G. RAND, AND G. PENNYCOOK (2023): “Reasoning about climate change,” *PNAS nexus*, 2, pgad100.
- BARRON, K. AND T. FRIES (2024): “Narrative persuasion,” Tech. rep., WZB Discussion Paper.
- BATTIGALLI, P. AND M. DUFWENBERG (2007): “Guilt in games,” *American Economic Review*, 97, 170–176.
- (2022): “Belief-dependent motivations and psychological game theory,” *Journal of Economic Literature*, 60, 833–882.

- BAYES, R. AND J. N. DRUCKMAN (2021): “Motivated reasoning and climate change,” *Current Opinion in Behavioral Sciences*, 42, 27–35.
- BAYES, R., J. N. DRUCKMAN, A. GOODS, AND D. C. MOLDEN (2020): “When and how different motives can drive motivated political reasoning,” *Political Psychology*, 41, 1031–1052.
- BICCHIERI, C. (2005): *The grammar of society: The nature and dynamics of social norms*, Cambridge University Press.
- BOLTON, G. E. AND A. OCKENFELS (2000): “ERC: A theory of equity, reciprocity, and competition,” *American Economic Review*, 91, 166–193.
- BRADY, W. J., J. A. WILLS, J. T. JOST, J. A. TUCKER, AND J. J. VAN BAVEL (2017): “Emotion shapes the diffusion of moralized content in social networks,” *Proceedings of the National Academy of Sciences*, 114, 7313–7318.
- BURSZTYN, L., G. EGOROV, AND S. FIORIN (2020): “From extreme to mainstream: The erosion of social norms,” *American Economic Review*, 110, 3522–3548.
- BURSZTYN, L., G. EGOROV, I. HAALAND, A. RAO, AND C. ROTH (2023): “Justifying dissent,” *The Quarterly Journal of Economics*, 138, 1403–1451.
- BURSZTYN, L. AND D. Y. YANG (2022): “Misperceptions about others,” *Annual Review of Economics*, 14, 425–452.
- BUSHELL, S., G. S. BUISSON, M. WORKMAN, AND T. COLLEY (2017): “Strategic narratives in climate change: Towards a unifying narrative to address the action gap on climate change,” *Energy Research & Social Science*, 28, 39–49.
- CAMERER, C. F., T.-H. HO, AND J.-K. CHONG (2004): “A cognitive hierarchy model of games,” *The Quarterly Journal of Economics*, 119, 861–898.
- CHAPMAN, D. A., B. LICKEL, AND E. M. MARKOWITZ (2017): “Reassessing emotion in climate change communication,” *Nature Climate Change*, 7, 850–852.
- CHEN, L. AND K. UNSWORTH (2019): “Cognitive complexity increases climate change belief,” *Journal of Environmental Psychology*, 65, 101316.
- CHERRY, C., C. VERFUERTH, AND C. DEMSKI (2024): “Discourses of climate inaction undermine public support for 1.5° C lifestyles,” *Global Environmental Change*, 87, 102875.
- COHEN, G. L. (2003): “Party over policy: The dominating impact of group influence on political beliefs,” *Journal of Personality and Social Psychology*, 85, 808.

- COSTA-GOMES, M. A. AND V. P. CRAWFORD (2006): “Cognition and behavior in two-person guessing games: An experimental study,” *American Economic Review*, 96, 1737–1768.
- DANA, J., R. A. WEBER, AND J. X. KUANG (2007): “Exploiting moral wiggle room: experiments demonstrating an illusory preference for fairness,” *Economic Theory*, 33, 67–80.
- DECHEZLEPRÊTRE, A., A. FABRE, T. KRUSE, B. PLANTEROSE, A. SANCHEZ CHICO, AND S. STANTCHEVA (2025): “Fighting climate change: International attitudes toward climate policies,” *American Economic Review*, 115, 1258–1300.
- DEL VICARIO, M., A. SCALA, G. CALDARELLI, H. E. STANLEY, AND W. QUATTROCIOCHI (2017): “Modeling confirmation bias and polarization,” *Scientific reports*, 7, 40391.
- DI TELLA, R., R. PEREZ-TRUGLIA, A. BABINO, AND M. SIGMAN (2015): “Conveniently upset: Avoiding altruism by distorting beliefs about others’ altruism,” *American Economic Review*, 105, 3416–3442.
- DIMANT, E., F. GALEOTTI, AND M. C. VILLEVAL (2024): “Motivated information acquisition and social norm formation,” *European Economic Review*, 104778.
- DIXON, G., C. CLARKE, J. JACQUET, D. T. EVENSEN, AND P. S. HART (2024): “The complexity of pluralistic ignorance in Republican climate change policy support in the United States,” *Communications Earth & Environment*, 5, 76.
- DOELL, K. C., P. PÄRNAMETS, E. A. HARRIS, L. M. HACKEL, AND J. J. VAN BAVEL (2021): “Understanding the effects of partisan identity on climate change,” *Current Opinion in Behavioral Sciences*, 42, 54–59.
- DRUCKMAN, J. N. AND M. C. MCGRATH (2019): “The evidence for motivated reasoning in climate change preference formation,” *Nature Climate Change*, 9, 111–119.
- DUFWENBERG, M. AND G. KIRCHSTEIGER (2004): “A theory of sequential reciprocity,” *Games and Economic Behavior*, 47, 268–298.
- EHRET, P. J., L. VAN BOVEN, AND D. K. SHERMAN (2018): “Partisan barriers to bipartisanship: Understanding climate policy polarization,” *Social Psychological and Personality Science*, 9, 308–318.
- EPLEY, N. AND T. GILOVICH (2016): “The mechanics of motivated reasoning,” *Journal of Economic Perspectives*, 30, 133–140.

- FALKENBERG, M., A. GALEAZZI, M. TORRICELLI, N. DI MARCO, F. LAROSA, M. SAS, A. MEKACHER, W. PEARCE, F. ZOLLO, W. QUATTROCIOCCI, ET AL. (2022): “Growing polarization around climate change on social media,” *Nature Climate Change*, 12, 1114–1121.
- FALLUCCHI, F., R. A. LUCCASEN III, AND T. L. TUROCY (2022): “The sophistication of conditional cooperators: Evidence from public goods games,” *Games and Economic Behavior*, 136, 31–62.
- FARROW, K., G. GROLLEAU, AND L. IBANEZ (2017): “Social norms and pro-environmental behavior: A review of the evidence,” *Ecological Economics*, 140, 1–13.
- FEHR, E. AND A. LEIBBRANDT (2011): “A field study on cooperativeness and impatience in the tragedy of the commons,” *Journal of Public Economics*, 95, 1144–1155.
- FEHR, E. AND K. M. SCHMIDT (1999): “A theory of fairness, competition, and cooperation,” *The Quarterly Journal of Economics*, 114, 817–868.
- FEHR, E. AND I. SCHURTENBERGER (2018): “Normative foundations of human cooperation,” *Nature human behaviour*, 2, 458–468.
- FEINBERG, M. AND R. WILLER (2013): “The moral roots of environmental attitudes,” *Psychological Science*, 24, 56–62.
- FERRARA, E., O. VAROL, C. DAVIS, F. MENCZER, AND A. FLAMMINI (2016): “The rise of social bots,” *Communications of the ACM*, 59, 96–104.
- FESTINGER, L. (1962): “Cognitive dissonance,” *Scientific American*, 207, 93–106.
- FIELDING, K. S., B. W. HEAD, W. LAFFAN, M. WESTERN, AND O. HOEGH-GULDBERG (2012): “Australian politicians’ beliefs about climate change: political partisanship and political ideology,” *Environmental Politics*, 21, 712–733.
- FISCHBACHER, U. AND S. GÄCHTER (2010): “Social preferences, beliefs, and the dynamics of free riding in public goods experiments,” *American Economic Review*, 100, 541–556.
- FISCHBACHER, U., S. GÄCHTER, AND E. FEHR (2001): “Are people conditionally cooperative? Evidence from a public goods experiment,” *Economics Letters*, 71, 397–404.
- FLØTTUM, K. AND Ø. GJERSTAD (2017): “Narratives in climate change discourse,” *Wiley Interdisciplinary Reviews: Climate Change*, 8, e429.

- FRANK, P., G. HENKEL, AND J. A. LYSGAARD (2024): “Between evidence and delusion—a scoping review of cognitive biases in Environmental and Sustainability Education,” *Environmental Education Research*, 30, 1477–1499.
- FUDENBERG, D. AND D. K. LEVINE (1993): “Self-confirming equilibrium,” *Econometrica: Journal of the Econometric Society*, 523–545.
- GEIGER, N. AND J. K. SWIM (2016): “Climate of silence: Pluralistic ignorance as a barrier to climate change discussion,” *Journal of Environmental Psychology*, 47, 79–90.
- GINO, F., M. I. NORTON, AND R. A. WEBER (2016): “Motivated Bayesians: Feeling moral while acting egoistically,” *Journal of Economic Perspectives*, 30, 189–212.
- GOLDBERG, M. H., S. VAN DER LINDEN, A. LEISEROWITZ, AND E. MAIBACH (2020): “Perceived social consensus can reduce ideological biases on climate change,” *Environment and Behavior*, 52, 495–517.
- GOLDBERG, M. H., S. VAN DER LINDEN, E. MAIBACH, AND A. LEISEROWITZ (2019): “Discussing global warming leads to greater acceptance of climate science,” *Proceedings of the National Academy of Sciences*, 116, 14804–14805.
- GOLMAN, R., D. HAGMANN, AND G. LOEWENSTEIN (2017): “Information avoidance,” *Journal of Economic Literature*, 55, 96–135.
- GOLMAN, R., G. LOEWENSTEIN, K. O. MOENE, AND L. ZARRI (2016): “The preference for belief consonance,” *Journal of Economic Perspectives*, 30, 165–188.
- GROSSMAN, Z. AND J. J. VAN DER WEELE (2017): “Self-image and willful ignorance in social decisions,” *Journal of the European Economic Association*, 15, 173–217.
- GUARINO, A. AND P. JEHIEL (2013): “Social learning with coarse inference,” *American Economic Journal: Microeconomics*, 5, 147–174.
- HART, P. S. AND E. C. NISBET (2012): “Boomerang effects in science communication: How motivated reasoning and identity cues amplify opinion polarization about climate mitigation policies,” *Communication Research*, 39, 701–723.
- HINKEL, J., D. MANGALAGIU, A. BISARO, AND J. D. TÀBARA (2020): “Transformative narratives for climate action,” .
- HORNSEY, M. J. (2008): “Social identity theory and self-categorization theory: A historical review,” *Social and Personality Psychology Compass*, 2, 204–222.
- HORNSEY, M. J. AND K. S. FIELDING (2020): “Understanding (and reducing) inaction on climate change,” *Social Issues and Policy Review*, 14, 3–35.

- JACHIMOWICZ, J. M., O. P. HAUSER, J. D. O'BRIEN, E. SHERMAN, AND A. D. GALINSKY (2018): "The critical role of second-order normative beliefs in predicting energy conservation," *Nature Human Behaviour*, 2, 757–764.
- JASNY, L., J. WAGGLE, AND D. R. FISHER (2015): "An empirical examination of echo chambers in US climate policy networks," *Nature Climate Change*, 5, 782–786.
- KAHAN, D. M., E. PETERS, M. WITTLIN, P. SLOVIC, L. L. OUELLETTE, D. BRAMAN, AND G. MANDEL (2012): "The polarizing impact of science literacy and numeracy on perceived climate change risks," *Nature Climate Change*, 2, 732–735.
- KATZ, D., F. H. ALLPORT, AND M. B. JENNESS (1931): "Students' attitudes; a report of the Syracuse University reaction study." .
- KUNDA, Z. (1990): "The case for motivated reasoning." *Psychological Bulletin*, 108, 480.
- LAMB, W. F., G. MATTIOLI, S. LEVI, J. T. ROBERTS, S. CAPSTICK, F. CREUTZIG, J. C. MINX, F. MÜLLER-HANSEN, T. CULHANE, AND J. K. STEINBERGER (2020): "Discourses of climate delay," *Global Sustainability*, 3, e17.
- LEE, S., H. T. VU, J. THAKER, M. VERNER, M. H. GOLDBERG, J. CARMAN, S. A. ROSENTHAL, AND A. LEISEROWITZ (2025): "Variations in climate change belief systems across 110 geographic areas," *Nature Climate Change*, 1–7.
- LEVISTON, Z., I. WALKER, AND S. MORWINSKI (2013): "Your opinion on climate change might not be as common as you think," *Nature Climate Change*, 3, 334–337.
- LYONS, B. AND A. HASELL (2024): "Communicating Republicans' level of support for climate policy briefly increases personal support in the United States," *Science Communication*, 46, 653–671.
- MA, Y., G. DIXON, AND J. D. HMIELOWSKI (2019): "Psychological reactance from reading basic facts on climate change: The role of prior views and political identification," *Environmental Communication*, 13, 71–86.
- MAIBACH, E., A. LEISEROWITZ, S. ROSENTHAL, C. ROSER-RENOUF, AND M. CUTLER (2016): "Is there a climate "spiral of silence" in America," *Yale Program on Climate Change Communication*.
- MARKS, G. AND N. MILLER (1987): "Ten years of research on the false-consensus effect: An empirical and theoretical review." *Psychological Bulletin*, 102, 72.
- MILDENBERGER, M. AND D. TINGLEY (2019): "Beliefs about climate beliefs: the importance of second-order opinions for climate politics," *British Journal of Political Science*, 49, 1279–1307.

- MILLER, D. T. AND C. MCFARLAND (1987): “Pluralistic ignorance: When similarity is interpreted as dissimilarity.” *Journal of Personality and social Psychology*, 53, 298.
- MOSER, S. C. (2010): “Communicating climate change: history, challenges, process and future directions,” *Wiley Interdisciplinary Reviews: Climate Change*, 1, 31–53.
- NICKERSON, R. S. (1998): “Confirmation bias: A ubiquitous phenomenon in many guises,” *Review of General Psychology*, 2, 175–220.
- NOELLE-NEUMANN, E. (1974): “The spiral of silence a theory of public opinion,” *Journal of Communication*, 24, 43–51.
- PRENTICE, D. A. AND D. T. MILLER (1996): “Pluralistic ignorance and the perpetuation of social norms by unwitting actors,” 28, 161–209.
- RABIN, M. (1993): “Incorporating fairness into game theory and economics,” *The American Economic Review*, 1281–1302.
- REYNAUD, A. AND B. OUVRARD (2024): “Re-calibrating beliefs about peers: Direct impacts and cross-learning effects in agriculture.” *TSE Working Paper*.
- RIMBAUD, C. AND A. SOLDÀ (2024): “Avoiding the cost of your conscience: belief dependent preferences and information acquisition,” *Experimental Economics*, 1–57.
- ROOS, M. AND M. RECCIUS (2024): “Narratives in economics,” *Journal of Economic Surveys*, 38, 303–341.
- ROSS, L., D. GREENE, AND P. HOUSE (1977): “The “false consensus effect”: An egocentric bias in social perception and attribution processes,” *Journal of Experimental Social Psychology*, 13, 279–301.
- RUSTAGI, D., S. ENGEL, AND M. KOSFELD (2010): “Conditional cooperation and costly monitoring explain success in forest commons management,” *Science*, 330, 961–965.
- SABHERWAL, A., A. R. PEARSON, AND G. SPARKMAN (2021): “Anger consensus messaging can enhance expectations for collective action and support for climate mitigation,” *Journal of Environmental Psychology*, 76, 101640.
- SANDERS, G. S. AND B. MULLEN (1983): “Accuracy in perceptions of consensus: Differential tendencies of people with majority and minority positions,” *European Journal of Social Psychology*, 13, 57–70.
- SCHULDT, J. P., Y. C. YUAN, Y. SONG, AND K. LIU (2019): “Beliefs about whose beliefs? Second-order beliefs and support for China’s coal-to-gas policy,” *Journal of Environmental Psychology*, 66, 101367.

- SHEERAN, P. AND T. L. WEBB (2016): “The intention–behavior gap,” *Social and personality psychology compass*, 10, 503–518.
- SHILLER, R. J. (2017): “Narrative economics,” *American Economic Review*, 107, 967–1004.
- SIMON, H. A. (1955): “A behavioral model of rational choice,” *The Quarterly Journal of Economics*, 99–118.
- SOKOLOSKI, R., E. M. MARKOWITZ, AND D. BIDWELL (2018): “Public estimates of support for offshore wind energy: False consensus, pluralistic ignorance, and partisan effects,” *Energy Policy*, 112, 45–55.
- SPARKMAN, G., N. GEIGER, AND E. U. WEBER (2022): “Americans experience a false social reality by underestimating popular climate policy support by nearly half,” *Nature Communications*, 13, 4779.
- SPARKMAN, G. AND G. M. WALTON (2017): “Dynamic norms promote sustainable behavior, even if it is counternormative,” *Psychological Science*, 28, 1663–1674.
- STURM, B., J. PEI, R. WANG, A. LÖSCHEL, AND Z. ZHAO (2019): “Conditional cooperation in case of a global public good—Experimental evidence from climate change mitigation in Beijing,” *China Economic Review*, 56, 101308.
- SULDOVSKY, B. (2017): “The information deficit model and climate change communication,” in *Oxford Research Encyclopedia of Climate Science*.
- TADDICKEN, M., S. KOHOUT, AND I. HOPPE (2019): “How aware are other nations of climate change? Analyzing Germans’ second-order climate change beliefs about Chinese, US American and German people,” *Environmental Communication*, 13, 1024–1040.
- TAJFEL, H., J. C. TURNER, W. G. AUSTIN, AND S. WORCHEL (1979): “An integrative theory of intergroup conflict,” *Organizational Identity: A reader*, 56, 9780203505984–16.
- TAVONI, A., A. DANNENBERG, G. KALLIS, AND A. LÖSCHEL (2011): “Inequality, communication, and the avoidance of disastrous climate change in a public goods game,” *Proceedings of the National Academy of Sciences*, 108, 11825–11829.
- TUCKETT, D. AND M. NIKOLIC (2017): “The role of conviction and narrative in decision-making under radical uncertainty,” *Theory & Psychology*, 27, 501–523.

TVERSKY, A. AND D. KAHNEMAN (1974): “Judgment under Uncertainty: Heuristics and Biases: Biases in judgments reveal some heuristics of thinking under uncertainty.” *Science*, 185, 1124–1131.

YANG, Y. AND J. E. HOBBS (2020): “The power of stories: Narratives and information framing effects in science communication,” *American Journal of Agricultural Economics*, 102, 1271–1296.