Communication and Social Norms: Experimental Evidence on Beliefs and Behavior

Friederike Johanna Reichel

Dissertation
Munich 2024

Communication and Social Norms: Experimental Evidence on Beliefs and Behavior

Inaugural-Dissertation
zur Erlangung des Grades
Doctor oeconomiae publicae (Dr. oec. publ.)
an der Ludwig-Maximilians-Universität München

2024
vorgelegt von
Friederike Johanna Reichel

Referent:	Prof. Dr. Klaus M. Schmidt
Koreferent:	Prof. Dr. Florian Englmaier
Promotionsabschlussberatung:	31. Januar 2024

Datum der mündlichen Prüfung: 26.01.2024

Namen der Berichterstatter: Klaus Schmidt, Florian Englmaier, Peter Schwardmann

Acknowledgements

Writing this thesis would not have been possible without the support of many people. First, I would like to thank my first supervisor, Klaus Schmidt, for his generous support, discipline, and sharp comments and for letting me explore my ideas freely. I would also like to thank Florian Englmaier, my second supervisor, for his constant encouragement and support and for always seeing the big picture. I thank Peter Schwardmann, my third supervisor and co-author, for providing me with many opportunities. Thanks to him, I started to work on research projects early on. Through our collaboration, I learned not to get discouraged when faced with setbacks. I thank him and Klaus Schmidt for the invaluable opportunity of a research stay at CMU Pittsburgh.

Over the years, many others have served as mentors for me. Yves Le Yaouanq is an inspiring example of academic rigor and integrity, and I greatly benefited from having him as a junior mentor. I thank Valeria Burdea for her encouragement and expertise and Andrej Woerner for providing excellent research infrastructure at MELESSA. I am grateful to Matthias Lang for our insightful discussions on the theoretical parts of this thesis. Simeon Schudy has left marks on the experimental design of all three essays. I thank him for providing me with a wonderful workplace at the Giselastraße. Everyone there made my life more enjoyable.

I thank my colleagues Anna, Carla, Christoph, Jae, Jinju, Lena, Marcel, Sebastian, Silvia, Svenja, and Timm for engaging in fruitful and enjoyable discussions during coffee breaks and lunchtime. I am grateful to have Peter Redler as a co-author; his econometric expertise greatly benefited our project. I thank Caroline Benkert, Manuela Firmin, and Sabine Wilhelm-Kauf for their assistance with administrative matters. I would like to acknowledge the financial support of the German Research Foundation (DFG) through the GRK 1928 and the CRC TRR 190. I am thankful to Carsten Eckel, Klaus Schmidt, and Georg Weizsäcker, the organizers of these programs, for providing excellent research conditions for doctoral students.

Lastly, I want to express my deepest gratitude to Tal, my family, and my friends for their support and patience throughout this journey.

Contents

Introduction

This thesis comprises three independent chapters. All three chapters share a common focus on communication and the dissemination of information. Using experimental evidence, each chapter sheds light on how information structures influence beliefs, social norms, and behavior.

The seminal works by Crawford and Sobel (1982) and later by Kamenica and Gentzkow (2011) have put communication and information design at the center of economic research. Successful communication can facilitate pursuing economic interests (Cooper et al., 1992; Bicchieri and Lev-On, 2007; Ginsburgh and Weber, 2020). Behavioral limitations and biases, however, threaten successful information transmission (Enke, 2020; Braghieri, 2021, 2023). Transmitting information on individual behavior by making behavior visible to others has been identified as a powerful tool to influence behavior (Bursztyn and Jensen, 2017). Letting people observe others' behavior allows them to update and correct their perceptions of the prevailing social norm. Perceptions of social norms, in turn, affect relevant behaviors, such as climate change attitudes (Andre et al., 2021) and labor force participation (Bursztyn et al., 2020).

Nevertheless, our understanding of how information structures shape beliefs, social norms, and relevant behaviors remains incomplete. Chapter 1 adds to this literature by investigating the impact of changing moral categories on behavior. It demonstrates that changes in the information structure, i.e., the moral language, can lead to misperceived social norms. Chapter 2 discusses how egocentric information processing can lead to information loss and belief polarization. Lastly, Chapter 3 tests whether lessons from the literature on peer effects and social pressure (Bursztyn and Jensen, 2017) can be leveraged to promote preventive health care behavior. In the following, I provide a brief overview of each chapter.

Chapter 1 investigates how the definition of moral categories affects moral behavior. Information structures can be a powerful tool to change behavior. Calling out more behaviors as "immoral" to induce better behaviors can be futile. Broadening a negative moral category normalizes falling into this category. In reaction to the diluted negative moral category, people may behave less morally —willing to bear the decreased reputational cost of being "immoral" (*diluting effect*). This paper conceptualizes the diluting effect and provides direct experimental evidence for it. The objective function of language designers (i.e., policymakers, activists) who shape moral categories governs the relative importance of the diluting effect. However, data from a second experiment suggest that language designers underestimate the diluting effect. The results have implications for designing moral categories and shed light on the origins of misperceived social norms.

Chapter 2, co-authored with Peter Schwardmann and Georg Weizsäcker, investigates the implications of egocentric information processing. When a speaker communicates with several audiences, the correct interpretation of the speaker's message requires an understanding of her incentives vis-a-vis all audiences. We hypothesize that, when confronted with this inference problem, some receivers neglect the extent to which a speaker also addresses other audiences. Such co-audience neglect will lead receivers whose incentives are aligned with the speaker's to be too trusting and receivers at odds with the speaker to be too skeptical. Co-audience neglect can lead to polarization in beliefs, even when audiences are exposed to identical information. We find suggestive experimental evidence for co-audience neglect in a simple disclosure game between a sender and two receivers.

Finally, Chapter 3, the product of joint work with Peter Redler, analyzes the potential for social choice architecture to increase take-up rates of breast cancer check-ups in a large sample of women in Germany. We provide causal evidence that the relative timing of check-up appointments among peers matters for participation: A woman is more likely to participate in breast cancer screening when her peers' appointments are scheduled shortly before her own. A simple intervention, however, shows that scheduling peers' appointments on the same day does not affect participation. We discuss possible mechanisms underlying the observed pattern of peer effects and highlight policy implications.

Chapter 1

Language and Social Norms: The Diluting Effect

1.1 Introduction

The last years have seen heated debates on political correctness (Morris, 2001; Braghieri, 2021) and a rise in right-wing populism (Guriev and Papaioannou, 2022). The discussions around political correctness show that our moral categories are in flux. Actions once deemed acceptable are now called “politically incorrect” or “environmentally unfriendly”. This has evoked the concern that these categories lose their moral force in the process (Miles, 1989). If people, as a consequence, stop caring about being called “environmentally unfriendly”, they might engage in less desirable behaviors than before. Despite the ubiquity of this argument in the public debate, a direct test of its merit is lacking.[1] If supported, this has direct implications for the design of moral language and social image based interventions.

Any identification of the effect of moral language on behavior faces the challenge of the interdependence between language and social norms (Galor et al., 2018; Tabellini, 2008; Ginsburgh and Weber, 2020). Since a change in the language can itself be the outcome of social change, any measurement of the effect using observational data is potentially biased.[2] To

[1] In 2022, a German federal minister, Claudia Roth, denounced calling climate activists “climate terrorists” since doing so would trivialize terror (tagesschau, 2023). In an article by The Economist (2021) titled ““Genocide” is the wrong word for the horrors China is inflicting on the Uyghurs” the authors warned against calling this crime a genocide because, ultimately, this strategy could turn labeling crimes as genocide into a less powerful tool to prevent such crimes. See Liao and Hansen (2023) for a discussion of more versions of this argument in the political debate.

[2] A recent example of social change manifesting in language is the revision of the ‘Racism’ entry by the Merriam-Webster dictionary at the request of an activist in the wake of the Black Lives Matter movement (Hauser, 2020). As an editor of the Merriam-Webster dictionary puts it, “Activism changes the language.” (Hauser, 2020).

overcome this challenge, I developed an experiment that exogenously imposes a moral language.[3]

Guided by a theoretical framework, this paper empirically studies (i) whether people engage in less moral behavior in response to a diluted negative moral category and (ii) how people perceive others to react to a broader negative moral category.

The theoretical framework characterizes the behavioral response to widening a negative moral category. It starts with the observation that our language does not allow arbitrary nuance. When a rich state space needs to be mapped into a more restricted signal space, the question of partitioning the action space naturally arises. I consider simple linguistic structures that partition the set of behaviors into two categories - a negative and a positive moral category. Following Bénabou and Tirole (2011), individuals differ in their inclination to behave morally but agree on the moral norm. Further, it is important to them how moral others consider them to be (for a review on how social image concerns affect behavior, see Bursztyn and Jensen, 2017). Crucially, others will base their judgment of an agent's morality not on the agent's action itself but rather on the moral category that the language associates with this action. To illustrate, imagine that Bob and Charlie can commute to work by car, bus, or bike. Assume that society cares about CO_2 emissions and considers commuting by car the least moral action and commuting by bike the most moral action.[4] The term "environmentally unfriendly" could either only refer to taking the car or to taking the bus as well. Throughout the paper, I refer to a linguistic structure that narrowly (broadly) defines the negative moral category as a narrow (broad) convention.[5] How does switching from a narrow convention to a broad convention affect behavior? Two patterns emerge. Take Bob to be a bus person: If he were not observed, he would take the bus. Under the narrow convention, taking the bus is "environmentally friendly", so he will do that. Since he cares about his social image, the broad convention induces him to take the bike, which now is the only "environmentally friendly" action. In response to the purified positive moral category of "environmentally friendly", Bob behaves more morally (purifying effect). Take Charlie, on the other side, to be a car person: If he were not observed, he would take the car. The narrow convention induces him to at least take the bus since this will allow him to escape the reputationally costly category of "environmentally unfriendly". The broad convention, however, dilutes the meaning of "environmentally unfriendly" by including the bus commute.

[3]As Djourelova (2023) shows, observational data from quasi-experiments is also well suited to overcome this challenge.

[4]Estimates from Our World in Data are 171 g Carbon dioxide-equivalents (CO_2eq) per passenger kilometer for a medium car (diesel) and 105 g CO_2eq per passenger kilometer for a bus. Estimates for cycling vary as they mainly depend on the cyclist's diet. Most sources suggest about 20 g CO_2eq per passenger kilometer.

[5]Due to restricting the analysis to two categories, a narrow negative moral category always implies a broad positive moral category.

Consequently, for Charlie, the convenience of going by car outweighs the reputational cost of being "environmentally unfriendly". In response to the diluted negative moral category of "environmentally unfriendly", Charlie behaves less morally (diluting effect). The theoretical analysis decomposes the total effect of broadening a negative moral category into the purifying effect and the diluting effect. I show that either of these effects can be dominant. If the diluting effect dominates the purifying effect, widening a negative moral category leads to less moral behavior on average.

Often, moral categories form in a decentralized way. The way people use them shapes their meaning.[6] Sometimes, however, policymakers or activists decide strategically how to communicate about behaviors. The model guides such language designers in trading off the diluting effect against the purifying effect. Both effects manifest in differences in the distribution of behavior across conventions. Therefore, it matters which value the language designer attributes to each action. This value function over actions may not be linear. There are many examples of concave or convex value functions. In the example of the commute to work, a language designer who mainly cared about CO_2 emissions would have a convex value function over the available actions. For an example of a concave value function, take dietary choices. As data reported by the Intergovernmental Panel On Climate Change (2022) show, switching from an omnivore diet to a vegetarian diet reduces CO_2 emissions by a lot more than switching from a vegetarian diet to a vegan diet. So, the shape of a language designer's value function may differ across settings, even when holding the overall objective (i.e., reducing CO_2 emissions) constant. Compared to a linear value function, a concave (convex) value function calls for a narrower (broader) convention.

The main experiment tests for the existence of the diluting effect. Using a within-subject design, I exogenously impose two linguistic structures that vary along the breadth of the negative moral category. Moral behavior, within the experiment, corresponds to protecting the environment: Participants who are students based in Munich, Germany, decide how much effort they want to provide to protect the environment.[7] The more real effort tasks they decided to work on, the more money is donated to moorland conservation, an effective natural climate solution. Using the strategy method, every subject decides how many real effort tasks she wants to solve for the environment in three treatments. In a baseline treatment, this choice is kept private. It reveals a subject's intrinsic motivation to protect the environment. Subsequently, subjects encounter the two remaining treatments in random order. In each, I exogenously impose a linguistic convention. A convention is fully characterized by the

[6] This thought goes back to Wittgenstein (1953) and features prominently in ordinary language philosophy and descriptive linguistics.

[7] Strong social norms around climate change in my sample led me to choose the frame of climate change.

threshold number of real effort tasks that a subject has to solve so as not to be announced to others as “not protecting the climate”. The narrow convention employs a lower threshold number of real effort tasks than the broad convention. After subjects have stated their three choices, one treatment is randomly selected. Subjects then have to solve the pre-specified number of real effort tasks. If one of the convention conditions is selected, participants are - based on their choice and the convention - announced to others as “not protecting the climate” or “protecting the climate”. Observing each subject’s behavior under both conventions is crucial to identifying the frequency of the diluting effect and the purifying effect.

The data are strong evidence of both effects predicted by the theoretical framework. A share of 27% of subjects’ behavior is consistent with the diluting effect. These subjects behave less morally once the negative moral category is widened.[8] An equally sized share of 25% of subjects’ behavior is consistent with the purifying effect. These subjects behave more morally once the negative moral category is widened.[9] Excluding subjects who do not react to a widening of the negative moral category, both effects account for 89% of reactions across the conventions.[10] The diluting effect does not dominate the purifying effect: On average, under both conventions, subjects provide the same level of effort to protect the environment. Thus, there are situations in which a language designer who cares solely about the average behavior that she induces would be indifferent between a broad negative moral category and a narrow negative moral category despite the diluting effect being at play. To test whether the shape of the value function indeed matters for the optimal convention, I apply convex and concave value functions to the experimental choice data and estimate the corresponding treatment effect of the broad convention. Consistent with the theoretical prediction, the narrow convention is preferable to the broad convention if the value function is concave. If the value function is convex, the opposite holds.

Unsurprisingly, the negative moral category is diluted when the convention switches from narrow to broad. Under the broad convention, more participants are “not protecting the climate” and, on average, these participants display a higher willingness to protect the environment. Participants, however, seem to underappreciate the extent of this dilution. Such perceptions should be the basis for language designers, i.e., policymakers, activists, and ordinary people, to use moral categories strategically. Effectively influencing others’ behavior

[8] More specifically, the diluting effect requires that someone is “protecting the climate” under the narrow convention and is lowering her effort under the broad convention to below the threshold of the narrow convention.

[9] More precisely, the purifying effect requires that someone is choosing an effort level smaller than the threshold of the broad convention as long as the narrow convention is in place but surpasses it, i.e., is “protecting the climate”, under the broad convention.

[10] The clear patterns of both effects cannot be accounted for by various pre-registered forms of decision noise as an alternative explanation.

through the categorization of their behavior requires an understanding of their reactions. Underestimating the diluting effect means underestimating the cost of ambitious moral categories, distorting the trade-off between narrow and broad negative moral categories. Such misperceptions would lead to suboptimally broad negative moral categories. To assess people's understanding of the purifying and diluting effect, I employ a second experiment.

In the Forecasting Experiment, a second set of subjects reports detailed and incentivized beliefs on how subjects participating in the original experiment reacted to the broad convention. Subjects significantly underestimate (overestimate) the diluting (purifying) effect in size and frequency, incorrectly believing that broadening a negative moral category will lead to more moral behavior. The apparent difficulty of understanding the effect of a change in the information structure on behavior gives rise to misperceived norms.

Conjointly, the theoretical and empirical analyses show that broadening negative moral categories indeed dilutes their meaning and evokes adverse reactions in parts of the population. The severity of the diluting effect depends on the language designer's value function. Language designers are, however, likely to underappreciate the diluting effect.

This paper contributes to several strands of the literature. First, it contributes to a growing literature on the economics of language (see Ginsburgh and Weber, 2020 for a review). Many investigations in this area have been purely theoretical, from the seminal work by Crawford and Sobel (1982) to the study of language in organizations (Crémer et al., 2007). The theoretical literature shows that, in an optimal language, the frequency of states should determine the breadth of labels associated with this state (Crémer et al., 2007; Dilme, 2018). I identify the shape of the language designer's objective function as another determinant of the coarseness of language. Experimental evidence on the economics of language is limited.[11] In recent work, Djourelova (2023) leverages a natural field experiment and finds that slanted language causally affects policy attitudes. Such natural field experiments are rare and typically do not allow for the type of within-person comparison required to investigate the diluting effect. I, therefore, develop a new experimental framework to study the effect of language on behavior. An open question is why language is used differently by different speakers (see Ginsburgh and Weber, 2020 for a discussion of linguistic diversity). Chauvin (2023) shows theoretically that heterogeneous use of language has negative welfare consequences. My results offer one explanation for why people use language differently. I document heterogeneous perceptions of how others behave under different conventions. This heterogeneity in perceptions gives speakers a motive to use different conventions, even if they agree on which behavior is desirable.

[11] Notable exceptions are Chen (2013) and Sutter et al. (2015) who use existing linguistic differences to explain economic behavior. Their findings support the Sapir–Whorf hypothesis, claiming that language has a causal effect on human behavior.

Second, this paper contributes to the literature on reputational or social image concerns. As the literature shows, social image concerns are a complex tool to wield. On the one side, switching on social image concerns by increasing the visibility of behavior can lead to higher adoption of the target behavior (Funk, 2010; Karing, 2023). Conversely, as Morris (2001) and Braghieri (2021) show, in the context of political correctness, social image concerns can lead to the loss of valuable information. Ali and Bénabou (2020) show that social image concerns may also hinder moral progress. I add to this nuanced picture of social image concerns by identifying the diluting effect. A policymaker who wants to nudge people towards some target behavior by defining good behavior and bad behavior needs to be aware that raising the standards of what is good inadvertently increases the acceptability of the bad category. Norm-based interventions (e.g., Gulesci et al., 2021; Karing, 2023) need to consider this countervailing force.

Third, this paper contributes to the literature on social norms, specifically misperceived social norms. Evidence is accumulating that people misperceive norms in the domains of health (Macchi, 2023), the labor market (Bursztyn et al., 2020), and climate change (Andre et al., 2021). Bursztyn and Yang (2022) provide a comprehensive review of the topic. Correcting these misperceptions can be a cheap and successful intervention to induce behavioral change (Bursztyn et al., 2020; Andre et al., 2021). Yet, little is known about the origin of misperceived norms. Braghieri (2021) offers an explanation by documenting that people's inferences from public utterances are often flawed. Publicly endorsed opinions that are subject to social pressure may, therefore, create a wrong impression of the prevailing norm. My results provide another but related microfoundation. Language is dynamic, and the apparent difficulty of people to map these changes into changes in behavior constitutes a source of misperceived norms. Related to the study of the effect of language on social norms is the study on the expressive function of law (Sunstein, 1996; Bénabou and Tirole, 2011; Lane et al., 2023). Lane et al. (2023) find compelling empirical evidence supporting the hypothesis that the law exerts a causal effect on social norms. Since laws and norms coevolve much like language and norms coevolve, Lane et al. (2023) also devise experiments for empirical identification. As language is a much less rigid codification of norms, I can directly manipulate language. While highly complementary, some key differences warrant the coexistence of both research endeavors. Few of us are lawmakers, but all of us are language designers. Thus, measuring how laypeople believe the law shapes social norms is of limited relevance, whereas measuring how laypeople believe language shapes social norms is of primary interest.

Section 1.2 sets up the motivating framework. Section 1.3 presents experimental evidence of the diluting effect, and Section 1.4 documents the underestimation of the diluting effect.

Section 1.5 concludes.

1.2 Motivating Framework

This section formally characterizes the behavioral response to a widening of a negative moral category. The diluting effect is part of this behavioral response and can even be dominant. The framework identifies the shape of a language designer's objective function as pivotal for the optimal breadth of a negative moral category. The theoretical insights inform the experimental strategy. All proofs are relegated to Appendix A.1.

1.2.1 Signaling under a Coarse Moral Language

The framework builds on the model developed by Bénabou and Tirole (2011) and adapted by Karing (2023). There is a continuum of agents with mass normalized to 1. Each agent has a type $v \in V \subset \mathbb{R}$ drawn according to a distribution $f(v)$ with full support. The set V is compact. An agent's type is her private information, but the distribution of types, $f(v)$, is common knowledge. An agent chooses an action $a \in \mathcal{A} = \mathbb{R}_0^+$. Actions are costly to the agent but beneficial for society. Lower actions are less moral than higher actions. The agent's costs of taking an action are described by the strictly increasing, continuous, twice differentiable, and strictly convex cost function $c : \mathcal{A} \to \mathbb{R}_0^+$ with $c(0) = 0$ and $\lim_{a\to\infty} c'(a) = \infty$. Actions generate a positive externality for society, as captured by the function $e : \mathcal{A} \to \mathbb{R}_0^+$ with $e(a) = a$. An agent's type v determines to what extent she internalizes the positive externality of her actions to society. Higher types, thus, have an inclination to take more moral actions. Agents care about their reputation; they are social image seeking. In particular, they care about the inferences others will make about their type (i.e., their innate morality v). Actions are perfectly observable in the original model by Bénabou and Tirole (2011). Agents then use these actions to make inferences on others' types. Often, however, we have to rely on imprecise signals of others' actions. We hear someone is "politically incorrect" or "environmentally unfriendly" without knowing precisely what they did. To capture these situations, I consider coarse linguistic structures. A convention, following the terminology by Chauvin (2023), partitions the action space into two moral categories: a bad moral category (b) and a good moral category (g). A convention, $m : \mathcal{A} \times \mathcal{A} \to \{g, b\}$ is fully characterized by a threshold action $r \in \mathcal{A}$

$$m(a,r) = \begin{cases} g & \text{if } a \geq r \\ b & \text{if } a < r. \end{cases}$$

The convention in place is common knowledge. That is, all agents agree on the definitions of moral categories.[12] All inferences on agents' types rely on the categorization provided by the convention. Agent v's utility is given by

$$U(a,v) = ve(a) - c(a) + \gamma E[v|m(a,r)],$$

where $\gamma \geq 0$ denotes the extent to which all agents are social image seeking. A useful concept is an agent's *natural action* (Frankel and Kartik, 2019). It is the action agent v would take absent any visibility or absent social image concerns (i.e. $\gamma = 0$),

$$a(v) = \arg\max_a ve(a) - c(a).$$

By the assumptions stated above, $a(v) = 0,\ \forall v \leq 0$. For all other agents, $v > 0$, the assumptions on $e(a)$ and $c(a)$ ensure a bijection between the type space V and the set of natural actions $\{a(v)|v > 0\}$. Knowing an agent's action would thus allow one to recover her type fully. Naturally, a convention introduces information loss. I restrict the analysis, however, to conventions where the moral categories reveal some information on agents' types.[13] Given a convention $m(a,r)$, I define the lowest type whose natural action qualifies her for the good moral category as v_r $(a(v_r) = r)$.

As Bénabou and Tirole (2011) show for the original version of the model, there exists a unique Perfect Bayesian Equilibrium of the signaling game induced above. The equilibrium is characterized by a cutoff type $\hat{v_r}$ who is indifferent between choosing her natural action $a(\hat{v_r})$ and taking action r.[14] In equilibrium, all agents of type $v \in [\hat{v_r}, v_r)$ take action r whereas all other agents of type $v \notin [\hat{v_r}, v_r)$ take their natural action (see Figure 1.1 for an illustration of equilibrium actions). Since the cutoff type $\hat{v_r}$ is known to everyone in equilibrium, the meaning of moral categories is commonly understood. Being "bad" is equivalent to being of type lower than $\hat{v_r}$.

[12]The class of conventions that I am considering coincides with the class of information structures that Karing (2023) is considering in her dynamic version of the model by Bénabou and Tirole (2011).

[13]I only consider conventions that ensure that the lowest (highest) type v_{min} (v_{max}) takes an action in the bad (good) moral category. Formally, I assume (i) $a(v_{max}) \geq r$, (ii) $a(v_{min}) < r$ and $U(r, v_{min}) - U(a(v_{min}), v_{min}) = v_{min}e(r) - c(r) - (v_{min}e(a(v_{min})) - c(a(v_{min}))) + \gamma\Delta(v_{min}) < 0$. See Equation 1.1 for a definition of Δ. These assumptions guarantee an interior solution and conveniently rule out out-of-equilibrium beliefs. In addition, I assume that $U(r,v) - U(a(v),v)$ is strictly monotone in v, i.e. $e(r) - e(a(v)) + \gamma\frac{\partial\Delta(v)}{\partial v} > 0$. This requires γ (i.e., social image concerns) not to be too large (in the case that $\frac{\partial\Delta(v)}{\partial v} < 0$).

[14]In the case of indifference, I assume that action r is taken. Specifically, the cutoff type $\hat{v_r}$ can be recovered by solving the fixed-point equation

$$U(r,\hat{v_r}) - U(a(\hat{v_r}),\hat{v_r}) = \hat{v_r}e(r) - c(r) - (\hat{v_r}e(a(\hat{v_r})) - c(a(\hat{v_r}))) + \gamma\Delta(\hat{v_r}) = 0, \tag{1.1}$$

where $\Delta(\hat{v_r}) = E(v|v \geq \hat{v_r}) - E(v|v < \hat{v_r})$.

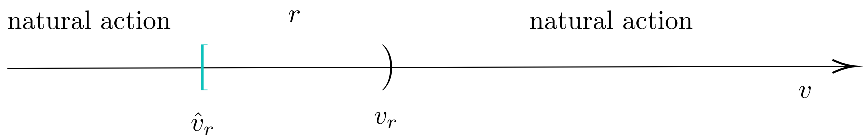


Figure 1.1: Equilibrium Actions under Coarse Moral Language

Notes: In equilibrium, all agents of type $v \in [\hat{v_r}, v_r)$ take action r whereas all other agents of type $v \notin [\hat{v_r}, v_r)$ take their natural action.

1.2.2 The Effects of Widening a Negative Moral Category

Since any convention induces a unique action profile, it is straightforward to compare conventions with respect to the behavior that they give rise to. In particular, I analyze the effect of widening the negative moral category.[15] Within the framework, I can study this effect by comparing moral behavior under a convention that narrowly defines the bad moral category with behavior that results from a convention that broadly defines the bad moral category. When comparing two conventions, I refer to the convention with the narrower (broader) negative moral category as the narrow (broad) convention. A narrow convention is characterized by a threshold action r_n that is lower than the threshold action r_b by which a broad convention partitions the action space into moral categories. Two opposing effects occur when switching from a narrow to a broad convention. The total effect of widening the negative moral category on behavior can be decomposed into what I call a purifying and a diluting effect,

$$\text{Total effect} = \underbrace{\int_{\hat{v}_{r_b}}^{v_{r_b}} (r_b - a(v)) f(v) dv}_{\text{Purifying effect}} - \underbrace{\int_{\hat{v}_{r_n}}^{v_{r_n}} (r_n - a(v)) f(v) dv}_{\text{Diluting effect}} . \tag{1.2}$$

On the one side, the increase in r from r_n to r_b purifies the good moral category, drawing in the types that were already close to choosing weakly more than the new threshold action $v \in [\hat{v_{r_b}}, v_{r_b})$. On the other side, the widening of the bad moral category dilutes its meaning, thereby drawing in types that were already close to choosing less than the previous threshold action $v \in [\hat{v_{r_n}}, v_{r_n})$. The direction of the total effect is ambiguous, with either of the two effects potentially dominating the other. The sign of the total effect is highly context-specific.[16] More generally, the decomposition shows that widening a negative moral category can only lead to less moral behavior on average if the diluting effect exists. Put differently,

[15] The analysis is related to the analysis of changing the threshold of a threshold bonus incentive (see, for example, Campos-Mercade and Wengström (2020)). A key difference is that the payoff for exceeding the threshold is fixed under a threshold bonus scheme. In the signaling model however, the social image "bonus" of exceeding threshold action r, $\Delta(\hat{v_r}) = E(v|v \geq \hat{v_r}) - E(v|v < \hat{v_r})$, varies with r.

[16] The ambiguity of the total effect has already been noted by Karing (2023). Its direction ultimately depends on the distribution of types. The variation in $\Delta(\hat{v_r}) = E(v|v \geq \hat{v_r}) - E(v|v < \hat{v_r})$ with r plays a key role (see Jewitt (2004) for notes on how $\Delta(\hat{v_r})$ varies with r for different distributions of v). A simple numerical example in Appendix A.1 illustrates the ambiguity of the total effect.

the diluting effect is a necessary condition for a total negative effect. This paper focuses on testing whether this condition is empirically plausible. If it is not, widening negative moral categories should not evoke much concern. If the data support the existence of the diluting effect, understanding contextual features that make it dominant or particularly costly should be the next step.

1.2.3 A Language Designer's Perspective

Often, moral categories form in a decentralized way. Sometimes, however, entities act as language designers. Politicians, activists, influencers, or organizations such as NGOs or governments define how broadly a negative moral category is used. Assume that a language designer can employ any convention of the abovementioned type. That is, she can freely choose a threshold action r that partitions the action space into a bad and a good moral category. Further, assume that the language designer's valuation of actions is described by the strictly increasing function $l : \mathcal{A} \to \mathbb{R}$. This value function $l(a)$ may or may not coincide with the externality function $e(a)$ that agents consider when taking their actions. Crucially, it may be non-linear. The language designer is assumed to know the unique equilibrium action profile that her choice of convention induces. This is reflected in her objective function $o : V \times \mathcal{A} \times \mathcal{A} \times \mathcal{A} \to \mathbb{R}$,

$$o(f(v), c(a), l(a), r) = \int_{v_{min}}^{\hat{v_r}} l(a(v))f(v)dv + \int_{\hat{v_r}}^{v_r} l(r)f(v)dv + \int_{v_r}^{v_{max}} l(a(v))f(v)dv. \quad (1.3)$$

The language designer applies her value function $l(a)$ to the equilibrium actions and aggregates over types. The choice between conventions is a choice between different distributions over actions. If the language designer's value function $l(a)$ is linear, the mean action induced by each convention is a sufficient statistic to identify the optimal threshold. This, however, no longer holds when the language designer's value function is concave or convex. If the language designer compares a narrow convention with a broad convention, a concave value function causes her to attach more weight to the diluting effect. Vice versa, a convex value function causes her to attach more weight to the purifying effect. This shifts the relative advantage of one convention over the other. To state this argument more formally, take a linear value function $l(a)$ and let $m(a, r^*)$ be the optimal convention that maximizes $o(f(v), c(a), l(a), r)$.[17] If $\tilde{l}(a)$ is concave, the optimal convention $m(a, \tilde{r}^*)$ must not define the negative moral category more broadly than the initial optimal convention (i.e. $\tilde{r}^* \leq r^*$). Conversely, if $\tilde{l}(a)$ is convex, the optimal convention $m(a, \tilde{r}^*)$ must not define the negative moral category more narrowly

[17] Since any convention $m(a, r)$ is fully characterized by its threshold action r, maximizing the objective function with respect to $m(a, r)$ is equivalent to maximizing it with respect to r.

than the initial optimal convention (i.e., $\tilde{r}^* \geq r^*$). The language designer's objective function, more specifically the shape of the value function, therefore affects the optimal breadth of the negative moral category.

1.3 Experimental Test for Diluting Effect

The purpose of the experiment is to test for the existence of the diluting effect. Measuring moral behavior under both a narrow negative moral category and a broad negative moral category allows me to decompose the average effect of widening a negative moral category on behavior. I explore how people perceive the shift in meaning that the negative moral category undergoes when widened.

1.3.1 Experimental Design

The experiment is designed to recreate the core features of the motivating framework. I employ two conventions that differ in their definition of the negative moral category. The narrow convention narrowly defines what it means to be "not protecting the climate" whereas the broad convention defines "not protecting the climate" more broadly.[18] In the experiment, I randomly vary the convention in place and measure the behavior that ensues. Since identifying the diluting effect and the purifying effect requires a within-subject design, every subject encounters both conventions.[19]

There are three experimental conditions: the Natural Action Condition, the Narrow Convention Condition, and the Broad Convention Condition. A subject takes an action a in each of the three experimental conditions: a^* denotes her action in the Natural Action Condition, a_n denotes her action in the Narrow Convention Condition, and a_b denotes her action in the Broad Convention Condition. Higher actions require more effort from the subject but contribute more to climate protection. The Natural Action Condition is devised to measure a subject's type, i.e., her natural inclination to provide effort for climate protection (*natural action*). The convention conditions introduce signaling motives. Each convention maps actions into the two categories of "not protecting the climate" and "protecting the climate". The mapping, however, slightly differs by convention. The Narrow Convention Condition distinguishes between the two categories at a lower threshold action than the Broad Convention Condition ($r_n < r_b$).

[18] Strong social norms around climate change in my sample led me to choose the frame of climate change (see Figure A.14).

[19] If I expect subjects to deviate from theory by choosing actions other than the threshold action or their natural action, the diluting and purifying effect cannot be recovered by the distribution of actions across conventions (i.e., by a between-subject design).

In particular, a subject's action corresponds to the number of real effort tasks she is willing to solve for climate protection. The real effort task is a variant of the counting zeros and ones tasks. Subjects need to count the number of frogs in a grid consisting of frog and broccoli emojis (see Figure A.20 for an illustration). For every real effort task subjects solve, I donate five eurocents to an environmental NGO (Bund Naturschutz in Bayern e.V.). The donations go towards moorland conservation, an effective natural climate solution. It is clear to subjects that the fewer tasks they choose to solve, the sooner they can leave the experiment. As every real effort task is longer by one line of emojis than the previous one, the marginal costs of taking a higher action are increasing. As intended, working on real effort tasks is perceived as moral yet individually costly: 84.71% of subjects view solving a real effort task as a good contribution to environmental protection, and 90.19% of subjects perceive the task as annoying.

A subject may choose not to work on any or up to 20 real effort tasks. The threshold actions are set to $r_n = 11$ and $r_b = 17$. Hence, compared to the Narrow Convention Condition, the Broad Convention Condition widens the meaning of "not protecting the climate" from solving less than 11 real effort tasks to solving less than 17 real effort tasks.

Using the strategy method, subjects state an action in the three experimental conditions, knowing that one condition will be drawn randomly. Once a condition is randomly selected, a subject must work on the number of real effort tasks she committed to under this condition. In addition, if any of the convention conditions is drawn at random, a subject's action is partially made public to the other participants. In particular, if the Narrow (Broad) Convention Condition is selected, and the subject's action a_n (a_b) is smaller than the threshold action r_n (r_b), she will be announced as "not protecting the climate", and as "protecting the climate" otherwise.

Figure 1.2 outlines a subject's path through the experiment. A subject joins her session on Zoom and turns on her camera for the entirety of the Zoom meeting. After consent, she reads the baseline instructions, familiarizes herself with the real effort task by solving two practice tasks, and answers four comprehension checks. She then states her natural action under the Natural Action Condition, knowing that her natural action will only be relevant with probability $\frac{1}{3}$. At that point, she is already aware that she will make the same choice again twice under slightly different circumstances (i.e., convention conditions).

Only after stating her natural action does she learn the details of the convention conditions. At first, the two convention conditions are explained in an abstract way. She learns that a convention is characterized by a threshold action r and that her action may be coarsely communicated to others. After passing two more comprehension checks, the subject encoun-

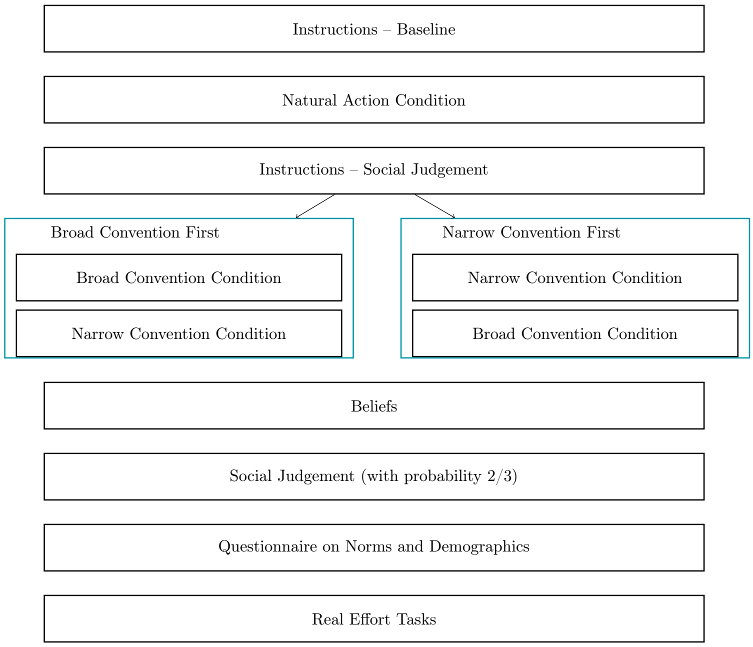


Figure 1.2: Experimental Design

ters the two convention conditions in random order.[20] In a given convention condition, she first learns the respective threshold action $r_{j\in\{n,b\}}$. Incentivized by the method developed in Krupka and Weber (2013), she indicates how socially appropriate she thinks others deem "not protecting the climate" (i.e. solving less than $r_{j\in\{n,b\}}$ real effort tasks) versus "protecting the climate" (i.e. solving weakly more than $r_{j\in\{n,b\}}$ real effort tasks). She is incentivized to state her belief on the share of subjects "not protecting the climate". After stating these beliefs, she takes her action $a_{j\in\{n,b\}}$.

Once she has encountered both convention conditions, she indicates her beliefs (i) on the average action under each condition and (ii) on the natural action of subjects who are "not protecting the climate" in each of the two convention conditions ($\hat{E}_{j\in\{n,b\}}[a^*|a_j < r_j]$). Then, one condition is randomly selected with equal probability. Public announcements are made should one of the two convention conditions be selected. The subject's video is spotlighted during the announcement, i.e., displayed in large to everyone in the session.[21] Hereafter, subjects leave the Zoom meeting and complete the experiment independently. They answer a few questions on social norms around climate change taken from Andre et al. (2021), fill out a brief demographics questionnaire, and solve the real effort tasks.

[20]The order is randomized conditional on gender to balance gender across treatments.
[21]This procedure was known to the subject when she took her actions. Reassuringly, 74% of participants believed the anticipated social judgment mattered for other participants' choices.

Procedural Details

A total of 268 subjects participated in 12 online sessions. The sessions took place on Zoom in January 2023. Participants are students based in Munich, Germany, registered with MELESSA (Munich Experimental Laboratory for Economic and Social Sciences) and recruited via the online recruitment system ORSEE (Greiner, 2004). The experiment was implemented with the survey software Qualtrics and pre-registered at the American Economic Association Registry for randomized control trials under trial number AEARCTR-0010740 (https://doi.org/10.1257/rct.10740-1.0). I exclude one person who participated twice in the experiment and 12 participants who did not arrive at the social judgment stage, i.e., participants for whom I do not observe all actions and beliefs. There are 2 participants who dropped out while working on the real effort tasks. Since I observe all outcome variables for them, I do not exclude them.[22] Hence, the main sample comprises 255 subjects. Subjects were, on average, 24 years old, and 51% of the sample is female. Less than 10% of participants reported personally knowing another participant in their session. The average time needed to complete the experiment (including the real effort tasks) was 74 minutes. Subjects earned, on average, 19 Euros, including a 6 Euros show-up fee, a flat payment of 12 Euros upon completion, and variable payment for the accuracy of their stated beliefs.

1.3.2 Experimental Evidence on the Diluting Effect

Three sets of results obtain. First, both effects proposed by theory - the diluting and purifying effect - are empirically relevant forces. Second, behavior is distributed very differently under the two conventions. This has implications for the optimal convention. Third, participants qualitatively understand the shift in the meaning of the negative moral category due to its widening but underestimate it quantitatively.

Diluting Effect and Purifying Effect

The data support the existence of the two effects proposed by the motivating framework: the diluting and the purifying effect. As pre-registered, I classify a subject's actions as consistent with the diluting effect if her action is weakly greater than r_n in the Narrow Convention Condition but smaller than r_n in the Broad Convention Condition. This is someone who is "protecting the climate" under the narrow convention but is "not protecting the climate"

[22] One concern might be that subjects choose an action higher than the threshold action in either of the two convention conditions to be announced as "protecting the climate" and then drop out in order not to bear the costs of their high action. I observe behavior consistent with this concern only for two participants. Both participated in a session where the Broad Convention Condition was randomly selected to be relevant. There were steep incentives in place to complete the experiment.

once the meaning of this category is diluted. Analogously, I classify a subject's actions as consistent with the purifying effect if her action is smaller than r_b in the Narrow Convention Condition and weakly greater than r_b in the Broad Convention Condition. Such a person may have been "protecting the climate" under the narrow convention or not but makes sure to be "protecting the climate" as this category is purified.

Figure 1.3 shows that 27.06% of subjects' behavior is consistent with the diluting effect and 25.49% of subjects' behavior is consistent with the purifying effect. Another 41.18% of subjects take the same action under the Narrow Convention Condition and the Broad Convention Condition. This leaves 6.28% of subjects' behavior unexplained. These sizable shares indicate clear patterns, pointing to the existence of both hypothesized effects.

However, decision noise or random responses could be an alternative explanation for the observed frequencies. I employ a binomial test against several pre-registered benchmark frequencies to account for decision noise. First, I compare the observed relative frequency of the diluting effect with the relative frequency that would arise if convention actions were randomly distributed around the natural action. For each individual, I simulate a narrow convention action $\hat{a}_n$ and a broad convention action $\hat{a}_b$ by independently adding uniformly distributed noise to their natural action a^*. I repeat this simulation $10,000$ times and record the relative frequency of the diluting effect. The average relative frequency of the diluting effect across all repetitions serves as the relative frequency of the diluting effect under the null hypothesis, p_0. Using the binomial test, I test whether the observed relative frequency of the diluting effect is significantly different from p_0. Figure A.4 displays the results of this testing procedure for the diluting effect. The first (second) benchmark is constructed by uniformly drawing noise from the set $\{-1, 0, 1\}$ ($\{-2, -1, 0, 1, 2\}$). I can thus reject the hypothesis that the observed share of the diluting effect is a mere result of this type of decision noise (p-value < 0.01). The same holds for the purifying effect when applying the analogous test (see Figure A.5, p-value < 0.01). Results using another benchmark confirm this finding. In this instance, I restrict the sample to subjects whose actions differ across the Narrow Convention Condition and the Broad Convention Condition. This time, I simulate a narrow convention action $\hat{a}_n$ and a broad convention action $\hat{a}_b$ with independent draws from the uniform distribution over the interval $[0, 20]$. I repeat this simulation $10,000$ times and record the relative frequency of the respective effect. The average relative frequency of the respective effect across all repetitions serves as the relative frequency of the effect under the null hypothesis, p_0. Using the binomial test, I find for both the diluting and the purifying effect that the observed relative frequency of each effect is significantly different from p_0 (p-value < 0.01). Since various forms of decision noise cannot explain the observed frequency of the diluting or the purifying effect, I interpret

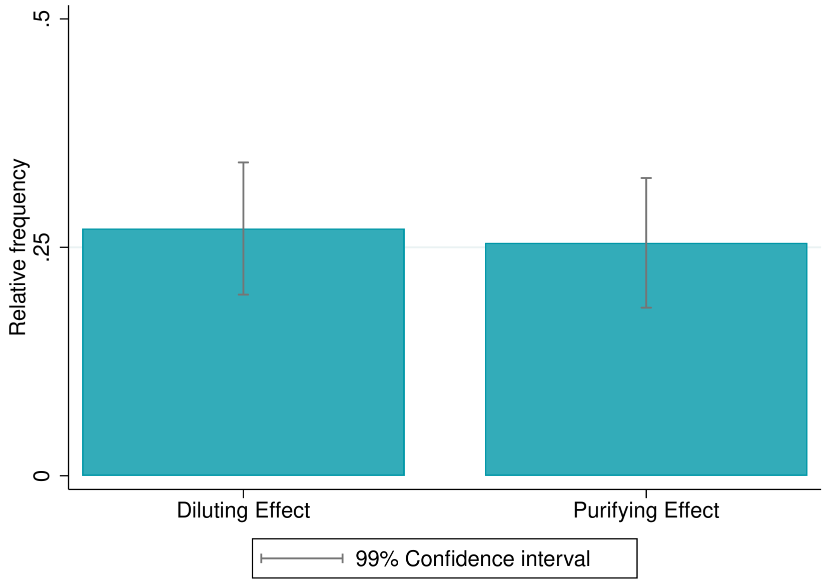


Figure 1.3: Diluting Effect and Purifying Effect

Notes: 27% (25%) of participants are behaving consistently with the diluting (purifying) effect. A subject who behaves consistently with the diluting effect chose to solve weakly more than 11 real effort tasks under the narrow convention and less than 11 real effort tasks under the broad convention (i.e., $a_n \geq r_n \wedge a_b < r_n$). A subject who behaves consistently with the purifying effect chose to solve less than 17 real effort tasks under the narrow convention and weakly more than 17 real effort tasks under the broad convention (i.e., $a_n < r_b \wedge a_b \geq r_b$).

the effects as systematic behavior in line with theory.

What predicts behavior in line with the diluting and purifying effect? A natural, and for that matter only, candidate proposed by the motivating framework is a person's type, as measured by her natural action. However, Figure A.6 in Appendix A.3 shows that this variable has limited explanatory power. The distributions of natural actions of those behaving consistently with the diluting effect and those acting consistently with the purifying effect are largely overlapping. At least, the average natural action of those behaving consistently with the diluting effect is significantly smaller than the average natural action of those acting consistently with the purifying effect (p-value < 0.01), suggesting some predictive power of the natural action. Consistent with the theoretical framework, a natural action below the threshold action of $r_n = 11$ is strongly correlated with the diluting effect (see Table A.7 in Appendix A.2). Analogously, a natural action between the two threshold actions, in the interval $[11, 17)$ is strongly correlated with the purifying effect (see Table A.8 in Appendix A.2). Self-reported social image concerns are correlated with the purifying effect but not the diluting effect. The few demographic variables elicited from the homogeneous student sample do not correlate with either of the effects. The order in which subjects encounter the conventions also does not change the relative frequencies of the effects. Beliefs on the net signaling benefit of the good category ($\Delta = E(v|a \geq r) - E(v|a < r)$) and how it changes across conventions are, for the most part, uncorrelated with either of the effects (see Tables A.7,A.8). Since these beliefs were not exogenously manipulated, it is impossible to conclude that beliefs do not matter for the effects. As Section 1.3.2 discusses, participants are generally aware of the shift in meaning that changes in the convention bring about.

Behavior under the Narrow and the Broad Convention

Not only are the diluting and purifying effects equally frequent (see Figure 1.3), but they are also similar in size. Whereas the purifying effect corresponds to an average increase in actions by 1.4, this gain is offset by an average decrease in actions by 1.5 due to the diluting effect. Since both effects absorb almost all changes in behavior in response to changing moral categories (see Figure A.2), it is unsurprising that the average action does not differ across conventions (see Table A.2). The breadth of the moral category does not affect the average effort participants provide to protect the environment. If anything, the mean action under the broad convention is insignificantly *lower* than the mean action under the narrow convention (p-value = 0.54). Recall that subjects encounter the conventions in a random order. Table A.2 column (2) restricts the analysis to the first action. This between-subject comparison indicates a significantly *higher* mean action under the broad convention. In combination with the negative and significant effect of the broad convention among the second actions in column (3), a differential dynamic of actions by the order in which subjects experience convention seems likely. I specifically test for order effects in column (4). The first coefficient indicates that subjects who encounter the narrow convention first choose a somewhat lower action under the broad convention (p-value = 0.15). Subjects who encountered the broad convention first chose, on average, a significantly higher action under the narrow convention than those who encountered the narrow convention first. Finally, the effect of the broad convention does not significantly differ by the order in which subjects encounter the treatments (p-value = 0.13). Table A.3 in Appendix A.2 adds control variables.[23] The results and their interpretation remain the same. The absence of clear order effects on the average action aligns with the finding that the frequency of both effects does not depend on the order of conventions either. Surprisingly, the mean action is not significantly higher in either of the convention conditions than under the Natural Action Condition (see Table A.4). This continues to hold even when pooling both convention conditions and despite 74% of subjects believing that the (partial) visibility and social judgment affect others' actions. In my setting, the null effect of visibility

[23]Table A.1 shows that all control variables are balanced across the order of treatments.

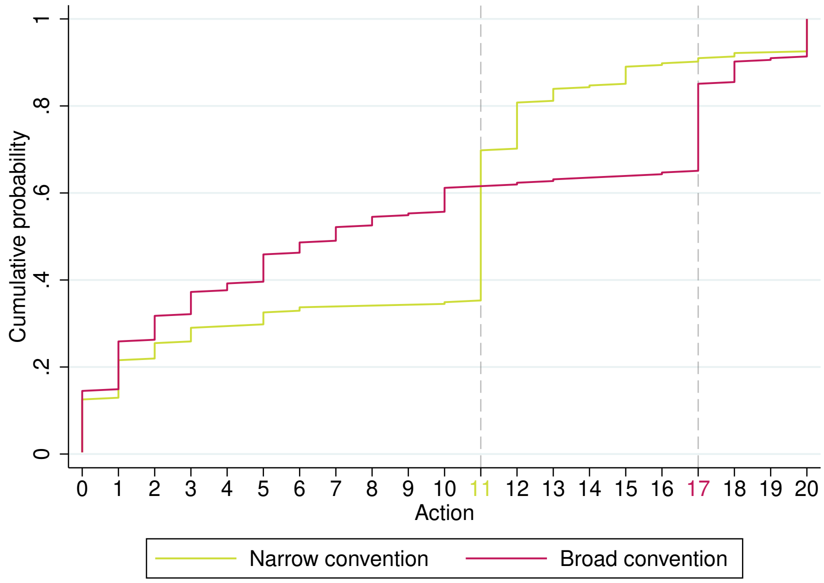


Figure 1.4: Cumulative Distribution Functions of Actions under Conventions

Notes: The threshold actions of both conventions are marked (i.e. $r_n = 11$, $r_b = 17$). Figure A.7 in Appendix A.3 adds the cumulative distribution function of the natural action to this graph. Figure A.8 in Appendix A.3 illustrates the bunching at the threshold actions.

on the mean action might be driven by order effects. The convention actions were elicited much later in the experiment than the natural action (see Figure 1.2).

Even though the mean, as an important moment of the distribution of actions, is unaffected by the breadth of the negative moral category, other moments are. For instance, the broad convention induces a significantly higher polarization in actions than the narrow convention (p-value of test of equal variance < 0.01). Figure 1.4 displays the cumulative distribution functions over actions by convention (p-value of Kolmogorov-Smirnov test < 0.01). Two observations are in line with the theoretical predictions. First, there is bunching at the threshold actions. Second, the cumulative distribution function induced by the narrow convention first-order stochastically dominates the cumulative distribution function induced by the broad convention for actions smaller than $r_n = 11$. The reverse is true for actions larger than $r_n = 11$ up until $r_b = 17$.

Relative Importance of the Diluting Effect

Which of the two conventions would a language designer prefer? The motivating framework provides some guidance. The mean action under both conventions is a sufficient statistic for a language designer who has a linear value function $l(a)$. Such a language designer, assuming that she had full access to the experimental data set, would be indifferent between the two

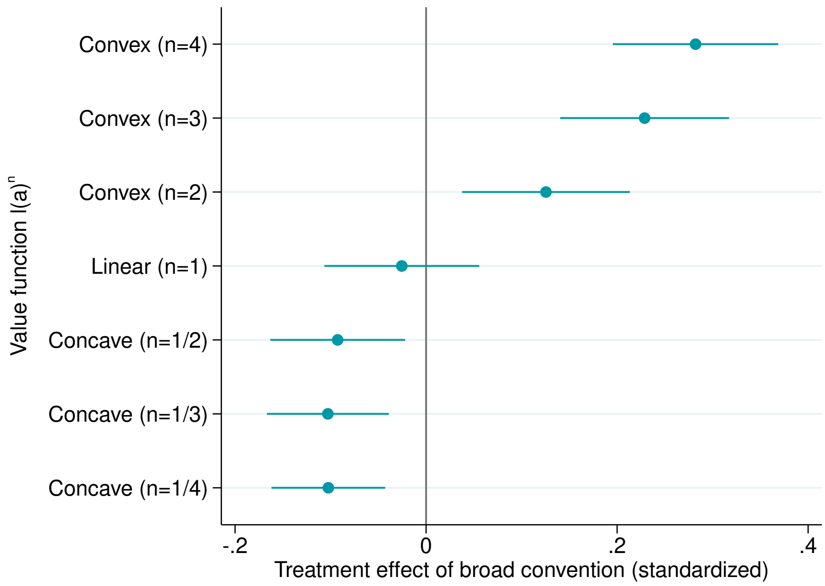


Figure 1.5: Effect of Broad Convention under Various Objective Functions

Notes: For each plotted coefficient, the respective value function $l(a)^n$ is applied to actions. The plotted coefficient represents the standardized treatment effect of the broad convention with 95% confidence intervals. The treatment effect plotted for the linear value function ($n = 1$) corresponds to the (standardized) average difference in actions between the two conventions.

conventions. More generally, a language designer who applies a non-linear value function to actions will consider the entire distribution. There are natural settings in which a language designer may have a concave or convex value function. Take the example of the commute to work. A car emits 171 g Carbon dioxide-equivalents (CO_2eq) per passenger kilometer, a bus emits 105 g CO_2eq per passenger kilometer, and a bike roughly 20 g CO_2eq per passenger kilometer.[24] Thus, a language designer who mainly cared about CO_2 emissions would have a convex value function over the available actions. Consider the case of dietary choices as an illustration of a concave value function. According to data from the Intergovernmental Panel On Climate Change (2022), transitioning from an omnivore diet to a vegetarian diet results in a greater reduction in CO_2 emissions compared to transitioning from a vegetarian diet to a vegan diet. Thus, the shape of a language designer's value function may vary between scenarios, even when the objective (i.e., reducing CO_2 emissions) is unchanged. To test the theoretical prediction that a concave value function makes the narrow convention more preferable, I first transform actions by applying a concave value function and then estimate the standardized treatment effect of the broad convention. Figure 1.5 plots the coefficients of standardized effects of the broad convention under various value functions.

[24] These numbers are based on estimates from Our World in Data.

Indeed, the negative effects of the broad convention under concave value functions support the comparative static of the model. Again, in line with theoretical predictions, convex value functions make the broad convention more preferable. The experimental design does not incorporate any value functions. In fact, the value function in this setting is up for debate. A linear value function seems natural since every real effort task generates a fixed donation. The increasing length of real effort tasks or convex opportunity costs could be cited as reasons for a concave value function. On the other side, training effects on the task would support a convex value function. The purpose of the exercise summarized in Figure 1.5 is to show that the shape of the language designer's objective function affects her choice of language. To this end, I use the experimental data but without taking a stance on which value function is the most plausible. Overall, the experimental data supports the guidance for a language designer provided by the theoretical framework.

Shifts in Meaning and Beliefs

Does broadening the negative moral category change the (perceived) meaning of the negative moral category? In line with the theory, the negative moral category is diluted. Subjects qualitatively understand this shift in meaning, but they underappreciate the extent to which the meaning of categories has changed.

The first measure of dilution is the share of subjects who take an action lower than the threshold action, i.e., the share of subjects who are "not protecting the climate". Under the narrow (broad) convention condition, this share corresponds to 35% (65%) (p-value of two-sided test of proportions < 0.01, see Table A.9). The second measure of dilution is the mean natural action conditional on an action lower than the threshold action ($E_{j\in\{n,b\}}[a^*|a_j < r_j]$).[25] I indeed find that the mean natural action of those who are "not protecting the climate" under the broad convention is significantly higher than the mean natural action of those who are "not protecting the climate" under the narrow convention (p-value of two-sided t-test < 0.01, see Table A.10).

Unsurprisingly, subjects correctly believe that fewer subjects are "not protecting the climate" under the narrow convention (43%) than under the broad convention (62%) (p-value of two-sided t-test < 0.01, see Table A.12). They also understand that the convention affects the inferences one can make about someone's type. Again consistent with the actual data, subjects believe that the mean natural action of those who are "not protecting the climate" under the broad convention is significantly higher than the mean natural action of those who

[25] The natural action is significantly correlated with a self-reported measure of environmental concern (p-value < 0.1). This supports the implicit assumption of my experimental design that the natural action can be interpreted as someone's type.

are “not protecting the climate" under the narrow convention (p-value of two-sided t-test < 0.01, see Table A.12).[26]

Comparing both measures of dilution (i.e., share of subjects in bad category and expected natural action conditional on bad category) with the perceived dilution suggests that subjects underestimate the extent to which the expansion of the bad moral category dilutes its meaning. First, on average, subjects believe there to be only 19 percentage points more subjects in the bad category once it is expanded, while actually, an additional 30 percentage points of subjects end up in the bad category (p-value of two-sided t-test < 0.01, see Figure A.11). Similarly, subjects believe on average that the mean natural action conditional on the bad category is higher by 1.8 tasks under the broad convention, whereas it is actually higher by 2.5 tasks (p-value of two-sided t-test < 0.01, see Figure A.11).

While subjects wrongly believe on average that either convention will cause significantly higher contributions than no convention, they correctly believe that the average action will not differ across the two conventions (see Figure A.13).

I document several correlations between own actions and beliefs on others' actions that align with the false consensus effect (Ross et al., 1977). First, a subject's belief on the average action strongly and significantly correlates with a subject's own action in each condition (p-value < 0.01 and correlation coefficient > 0.7 in all three cases). Second, subjects who contribute more under the broad convention than under the narrow convention believe that other subjects react similarly to the broad convention: The difference in actions under conventions and the difference in beliefs on the average contribution under each convention are significantly correlated (p-value < 0.01). Third, those whose actions are consistent with the diluting (purifying) effect are more likely to believe that the broad convention generates *lower* (*higher*) average contributions than the narrow convention (p-value in both cases < 0.01). Lastly, a subject's chosen language, i.e., the threshold action believed to generate the most contributions, is positively correlated with a subject's natural action (p-value < 0.01).[27] Surprisingly, only 45% of subjects choose a threshold action weakly smaller than their natural action, i.e., a language that would put themselves in the good moral category when simply choosing their individually preferred action.

[26] Table A.13 reports qualitatively similar results regarding the social appropriateness of either category across conventions. First, subjects believe the good category to be always more socially appropriate than the bad category. Second, both categories are believed to be more socially appropriate under the broad convention.

[27] In contrast to all other measures reported in this section, the choice of language was not incentivized.

1.4 Forecasting Experiment

It remains an open question whether language designers take the diluting effect into account. The Forecasting Experiment takes the first step toward answering this question. I measure how people perceive others to react to a widening of a negative moral category. I use the data set of the main experiment to benchmark and incentivize predictions. Participants of the Forecasting Experiment are drawn from the same student population as participants of the main study. If participants indeed underestimate the diluting effect, this would be a first indication that more influential language designers, such as NGOs, activists, or politicians, may misperceive the impact of using categories like "environmentally unfriendly" or "politically incorrect" more broadly.

1.4.1 Experimental Design

First, subjects are informed about the original experiment. They learn about the choices available to subjects in the original experiment, the impact of these choices (i.e., moorland conservation), the social judgment, and the two possible definitions of "not protecting the climate". They complete the same practice real effort task that was part of the original study. The wording is kept as close as possible to the initial study. Unbeknownst to them, subjects are randomly assigned to forecast behavior either under the broad convention (Predict Broad Convention condition) or the narrow convention (Predict Narrow Convention condition). For instance, when assigned to the Predict Broad Convention condition, subjects are informed about the distribution of behavior that occurred under the narrow convention in the original experiment. Their task is to predict how subjects behaved under the broad convention in the original experiment. The elicitation of these predictions is designed to be unintrusive to avoid prime subjects on either the diluting or purifying effect. Conditional on each possible action in the original experiment (solving 0 tasks, solving 1 task,..., solving 20 tasks), I elicit the predicted distribution of behavior under the broad convention condition. For instance, 8 participants chose to solve five tasks under the narrow convention in the original experiment. By adjusting bars, subjects can express their beliefs on how many tasks these participants chose to solve under the broad convention (see Figure A.21 for an illustration). Thus, every subject states 21 of these conditional distributions. From these rich data, I can recover (i) the perceived size and frequency of the diluting effect, (ii) the perceived size and frequency of the purifying effect, and (iii) the belief on the average action taken under the convention that is to be predicted. Subjects familiarize themselves with the elicitation and have to

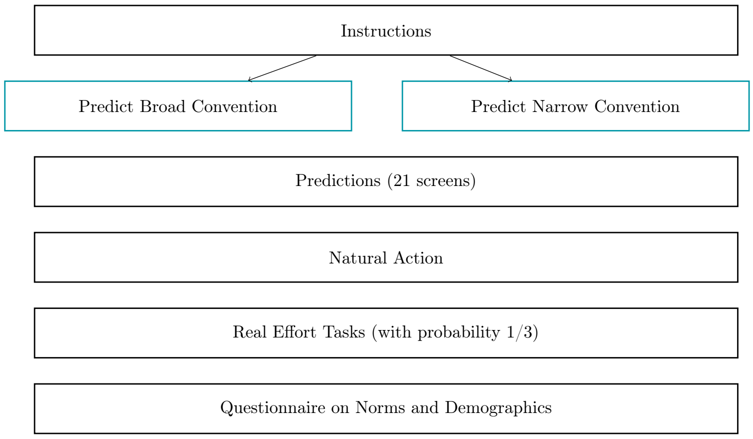


Figure 1.6: Forecasting Experiment - Design

correctly answer 5 comprehension questions.[28] Predictions are incentivized.[29] After stating all predictions, subjects indicate how many real effort tasks they want to solve without being observed (i.e., their natural action). The same incentives are in place as in the original study (i.e., every real effort task generates a donation of 5 eurocents for moorland conservation). With probability $\frac{1}{3}$, their choice is implemented, and they have to solve the number of tasks they committed to. At the end of the experiment, I elicit norms around climate change (Andre et al., 2021), strategic sophistication, and demographics. Figure 1.6 illustrates the experimental design.

The subjects of the Forecasting Experiment are deliberately drawn from the same population as the subjects from the original study. This reduces the likelihood of measuring misperceptions due to a lack of understanding of the social and cultural context in which the original study was conducted. All participants are students based in Munich and registered with the LMU Social Sciences Laboratory. Yet, as Table A.15 shows, baseline differences exist. While the average age is the same, the Forecasting Experiment has a significantly higher share of female participants. Due to no gender differences in the original study and the limitations of the subject pool, I did not enforce gender balance. As I discuss in the notes of Table A.15, the measures of environmental attitudes (including the natural action) vary across the two samples. Some of these differences are statistically significant but are generally small in size and possibly affected by the preceding elements of the experiments.

[28] 176 out of 198 participants stated that the instructions were easy to understand.

[29] The incentive scheme works as follows. For each prediction, one bar is randomly selected. Participants can earn between 0 and 100 points per prediction. The number of points earned for each prediction follows the formula $y = \max\{100 - (100\frac{x}{n} - 100\frac{g}{n})^2, 0\}$, where x represents the true value, g represents the predicted value, and n represents the number of participants over whom the prediction is made. The average of points is taken across all predictions and is used as the likelihood of winning the bonus of 5 Euros.

Procedural Details

A total of 208 subjects participated in the experiment. The main sample comprises the 198 subjects who completed the entire experiment. The data was collected in 2 waves in May and June 2023. Participants were students based in Munich, Germany, registered with MELESSA (Munich Experimental Laboratory for Economic and Social Sciences) and recruited via the online recruitment system ORSEE (Greiner, 2004). The experiment was pre-registered at the AEA Registry and implemented with the survey software Qualtrics. Subjects were, on average, 24 years old, and 72% of the sample was female. The median time needed to complete the experiment (including the real effort tasks) was 32 minutes, and subjects earned on average 12 Euros, including a 6 Euros show-up fee and a flat payment of 2 Euros upon completion.[30]

1.4.2 Underestimation of Diluting Effect

On average, participants significantly underestimate the frequency of the diluting effect (see Figure 1.7). At the same time, they significantly overestimate the frequency of the purifying effect on average.

These average misperceptions are not driven by outliers (see Figure A.16). A majority of 76% of subjects underestimate the frequency of the diluting effect and a majority of 67% of subjects overestimate the frequency of the purifying effect (see Table A.16).[31] Interestingly, a fifth of participants do not think about the diluting effect at all. This is only true for 7% of participants with respect to the purifying effect. The purifying effect may come to mind more easily.[32]

The data from the Forecasting Experiment confirm the previous finding that people underestimate the extent to which a category's meaning is diluted when it is widened. Again, subjects underestimate the additional share of subjects that are "not protecting the climate" once the broad convention is in place (see Figure A.19).

People do not only underestimate the diluting effect in terms of frequency, but they also do so in terms of size. As Table A.16 shows, the diluting effect is responsible for 1.51 tasks less per person but is believed to be responsible only for 1.26 tasks less per person. Qualitatively, the opposite is true for the purifying effect. Consistent with the misperceptions of both effects,

[30] The average time needed to complete the experiment was 117 minutes. The distribution of time needed is fairly right-skewed. It seems that a few people used the opportunity to pause the experiment and return to it later.

[31] In fact, more than half of the subjects (57%) underestimate the frequency of the diluting effect and at the same time overestimate the frequency of the purifying effect (see Figure A.17).

[32] The elicitation method required subjects to think of either of the effects by themselves. In particular, subjects were not given a definition of either of the effects. Once people's attention is drawn to the diluting effect, they may find it very plausible.

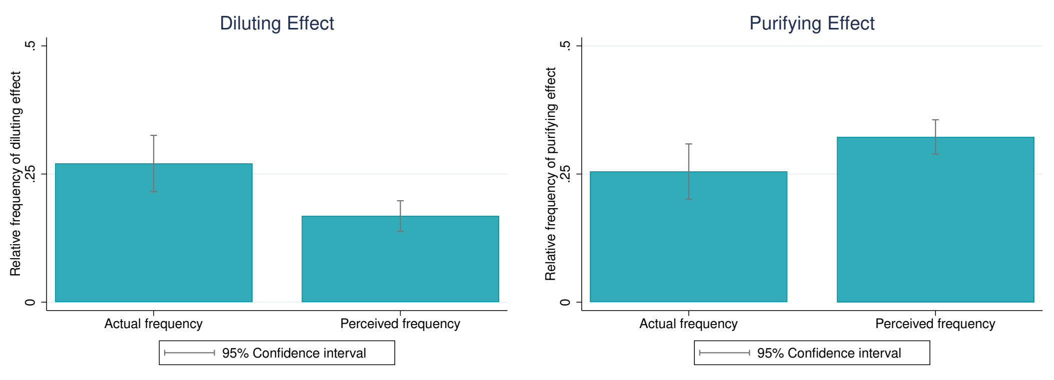


Figure 1.7: Perceptions of Diluting and Purifying Effect

Notes: The left panel displays the actual frequency (27%) and the perceived frequency (17%) of the diluting effect. The right panel displays the actual frequency (25%) and the perceived frequency (32%) of the purifying effect. The bars represent 99% confidence intervals.

subjects wrongly believe that, on average, the broad convention leads to more moral behavior than the narrow convention (see Figure A.18).[33] Thus, changing the language can be the source of misperceived social norms.

The misperceptions of the effects are significantly more pronounced among subjects who were randomly assigned to predict behavior under the broad convention (see Figure A.15 and Table A.18). Subjects who were informed about behavior under the broad convention and predicted behavior under the narrow convention do not, on average, misperceive the frequency of either of the effects. A randomization failure is unlikely to account for this stark difference (see Table A.14). Naturally, the elicitation screens differed across the two treatments in the Forecasting Experiment. In principle, random responses could lead to different inferred perceptions. Uniformly randomizing across answers under the Predict Narrow Convention condition would generate a higher (lower) perceived frequency of the diluting (purifying) effect than randomizing answers under the Predict Broad Convention condition. The data exhibit the opposite. Thus, the data suggest that the difficulty of correctly assessing people's reactions to a change in moral categories depends on the direction of change. Widening a negative moral category leads to more prevalent misperceptions than narrowing a negative moral category.

1.5 Discussion

Theoretically and empirically, this paper decomposes the effect of broadening a negative moral category. In response to the diluted meaning of the negative moral category, some people

[33] These beliefs were not directly stated but are recovered from the elicited conditional distributions.

behave less morally. Others, motivated by the purified meaning of the positive moral category, engage in better behaviors. The experimental data corroborate these two opposing effects, implying that the trade-off between these forces is real.

The insights are not confined to the realm of moral terms or categories but apply more generally to the design of information structures or labels that bestow reputation.[34] Consider education - a prominent example in economic signaling theory (Spence, 1973). A reform that raises the requirements to obtain the high school diploma certainly increases the signaling value of the diploma, pushing some students to perform better. But it also normalizes failing to get the certificate, primarily demotivating lower-ability students. How a policymaker should resolve this trade-off depends on her objective function. If she believes in relatively high returns to education for low-ability students, the diluting effect weighs more heavily, making the reform unattractive. Otherwise, if the policymaker perceives the returns to education for high-ability students as particularly high, the purifying effect becomes more significant, making the reform more appealing. Such reasoning about the optimal information structure requires an awareness of both effects. The second experiment presented in this paper indicates that the diluting effect is underestimated, giving rise to a distorted evaluation of the costs of ambitious moral categories or social labels. Such misperceptions present a threat to effectively using reputational concerns to induce target behaviors. Yet, it remains an open question whether these misperceptions translate into a preference for sub-optimally defined categories.

[34] There are two recent examples of broadening the negative category/narrowing the positive category. In 2021, the EU energy efficiency labels were adjusted such that, at least at the beginning, almost no electronic products qualify for the best category. In 2023, a German non-profit environmental and consumer protection association won a lawsuit against a large convenience store chain (dm). The association accused the chain of greenwashing by using its climate-neutral label too broadly.

Chapter 2

Co-audience Neglect

with Peter Schwardmann and Georg Weizsäcker

2.1 Introduction

An understanding of the context of public statements is necessary for a correct interpretation of those statements. The composition of the audience is important context: the same statement may mean different things depending on who else is listening, as the speaker may have different persuasion goals for different listeners. Yet, understanding the different goals vis-a-vis all listeners is difficult. In line with the psychological literature on egocentrism (Gilovich and Savitsky, 1999; Gilovich et al., 2000; Babcock et al., 1995), we hypothesize that, when faced with this difficulty, people tend to neglect other receivers when evaluating the veracity or informativeness of public statements. A person suffering from co-audience neglect feels a speaker is mainly addressing them, not the other listeners.

Consider a secretary of defense who learns through an internal report that her army is in disastrous shape. She might disclose this private knowledge or not. In deciding whether to make the report public, she will weigh her interests towards her audiences – the public and foreign governments - against each other. Assume that she wants the public to be correctly informed but foreign governments to believe that her army is in good shape regardless of its actual condition. Now, the public and the foreign governments will interpret the absence of such a public report very differently if they ignore the other party's existence. The public will conclude from the lack of a publicly available report that no such report has ever been written. At the same time, foreign governments will remain skeptical and consider it likely that a report is being withheld. Markets for banned goods, such as drugs, are another context in which co-audience neglect, if existent, would affect inferences and communication. A seller of a banned good faces conflicting interests vis-a-vis potential customers and law enforcers. If the seller has the banned good, she has an interest in conveying this to her customers, but at the same time, she wants to hide this from law enforcers. Co-audience neglecting law enforcers would incorrectly interpret the seller staying silent as completely uninformative as to whether she holds the forbidden good or not. Note that these examples share one feature: The sender's message is public. If the sender was able to send private messages to parts of the audience, co-audience neglect would not bias receivers' inferences.[1]

In general, co-audience neglect leads those whose incentives are aligned with the speaker to be too trusting and those who are at odds with the speaker to be too skeptical. As a result, it can be a source of polarization. Importantly, it can lead to polarization even when there is a single source of information: everyone listens to the same information, but the skeptics remain skeptical as the believers become more believing. It suggests that exposing everyone to the same information might not be sufficient to achieve depolarization. Recent evidence documents persistent or even increasing polarization in response to a common signal in contexts of public communication ranging from the Weimar Republic (Adena et al., 2015) to the Russo-Ukrainian war (Peisakhin and Rozenas, 2018) and Turkey's 2017 constitutional referendum (Baysan, 2022).[2]

In this paper, we deliver an experimental test of co-audience neglect. In a controlled experiment, we can investigate asymmetric updating with a single information source. This is hard to obtain in observational data, where selective information processing can lead to polarization. In the experiment, all listeners have identical information (in fact, identical instructions throughout the experiment), and we can investigate whether the preferences being aligned influences the inferences that listeners draw.

To bring structure to our experiment and our analysis of the bias, we first model a simple disclosure game. A sender has private information on the state of the world (high or low) and can send a public message to two receivers. The sender wants one of the receivers, the aligned receiver, to know the state of the world, whereas she wants the other receiver, the non-aligned receiver, to think that the state is low - regardless of the actual state. The sender's only means of achieving her persuasion goals is through her choice of the public message. If the state is high, then she can either disclose the state by sending a verifiable message or not disclose the state by sending an empty message. If the state is low, then she has no choice but to send an empty message. Both receivers' payoffs are increasing in the accuracy of their guess about

[1] We believe that situations of public communication are common and becoming more so, given technological and social changes such as the rise of social media or machine translation services.

[2] For theoretical explanations for persistent polarization despite a common signal, see Piketty (1995), Acemoglu et al. (2016), Gentzkow et al. (2018).

the state of the world.

The two receivers' beliefs are identical in the Bayesian Nash equilibrium of the game. However, a receiver with co-audience neglect ignores the other receiver's presence and their impact on the sender's choice of message. Therefore, under co-audience neglect, interpretations of the empty message differ. Because the aligned receiver expects the sender always to disclose the high state of the world, he will consider the high state unlikely when seeing the empty message. The non-aligned receiver does not think that the sender will ever disclose the high state of the world, so he will be maximally skeptical of the empty message and stick to his prior.

In our experiment, we let participants play the disclosure game between one sender and two receivers in several online sessions. Participants were recruited from a conventional student subject pool. Our main treatment variation is the random assignment of participants to receiver types – aligned versus non-aligned. A second treatment varies how much weight the sender attaches to her persuasion goals vis-a-vis each receiver. In the high alignment condition, she attaches more weight to the aligned receiver's beliefs such that the game's equilibrium predicts that she discloses her information in this condition and does the opposite in the non-aligned condition. We used the metaphor of gold mining to make the strategic incentives easy to understand. Each participant is assigned to a fixed role: gold miner (sender), partner (aligned receiver), and bandit (non-aligned receiver). The gold miner wants the partner to know whether she found gold or not (state of the world) and the bandit to believe that she found no gold. Participants play the disclosure game eight times in randomly determined groups. As usual in models with rational agents, the equilibrium also predicts that both receivers update their beliefs in the same way. The prediction under co-audience neglect, in contrast, is that the receivers interpret the empty message quite differently: upon seeing an empty message, aligned receivers' beliefs in a high state are lower than those of the non-aligned receivers in both conditions of the game.

We find that aligned receivers have more trust in the sender's message than non-aligned receivers. Compared to non-aligned receivers, aligned receivers believe on average that it is 4.5 percentage points less likely that the state is high when seeing the empty message (p-value = 0.06). Surprisingly, exposure to a sender's communication strategy that can only be rationalized by considering other listeners does not seem to be a channel through which co-audience neglect is alleviated. The average treatment effect is driven by a subset of participants, which is robust to different classifications of co-audience neglecting individuals. The measured difference in beliefs is robust to including unbalanced characteristics across treatments and excluding irrational or inattentive participants beyond our initial screening. It is, however, not robust to including several standard control variables at once, such as gender and age, that were balanced across treatments. It is also not robust to restricting the sample strictly to the pre-registered sample size (p-value = 0.13).[3] Senders expect co-audience neglect, as measured by their stated beliefs about receiver beliefs. We also find that senders react strongly to the alignment condition with 84% of truthful disclosure decisions in the high alignment condition compared to 31% percent in the low alignment condition.

Since co-audience neglect is based on a misperception of the sender's actual incentives, there is a risk that an experiment that obfuscates certain pieces of information artificially generates the bias under investigation. We conduct a strong experimental test of the bias by taking several measures to simplify receivers' inference. Most importantly, perhaps, the two audience members read precisely the same set of instructions, except for their role assignment. This minimizes the scope for an asymmetric understanding of the message. We study a simple game with only two receivers and a unique Bayesian Nash equilibrium. We provided easily accessible summaries of all instructions for the subjects to see as they click through the pages. Participants had to pass a set of nontrivial attention checks and answer comprehension questions about the instructions. Finally, by implementing transparent within-subject variation in the alignment of the sender's and receivers' incentives (high alignment versus low alignment condition), we repeatedly made salient the sender's incentives and how they relate to both receivers.

Co-audience neglect is an instance of context neglect (Eyster et al., 2015; Enke and Zimmermann, 2019; Enke, 2020). In the case of co-audience neglect, it is the other people listening – the co-audience – who are the part of the strategic environment that is neglected. Battaglini and Makarov (2014) suggest that receivers in a cheap talk game are insufficiently sensitive to the alignment of incentives between the sender and other receivers. Despite this established evidence, we are, to our knowledge, the first to deliver a direct test of the co-audience hypothesis and to show theoretically the implications of the bias.

Our results can also be interpreted through the lens of egocentrism. Conlon et al. (2021) find that information collected by people other than oneself is heavily discounted despite aligned interests and familiarity among those communicating, leading to insufficient learning. Similarly, Hyde (2021) shows that for disaggregated data, people overreact to signals pertaining to the category to which they belong. In our case, egocentrism manifests in suppressing the impact of the co-audience on the information shared.

The next Section is a theoretical analysis of the bias. Section 2.3 describes our experimental design, and Section 2.4 contains our results. Section 2.5 concludes.

[3] For a discussion of deviations from the pre-analysis plan, see Section 2.3.

2.2 A Disclosure Game with Co-audience Neglect

We set up a simple communication game to formalize the effect of co-audience neglect on beliefs and strategies. The analysis also yields testable predictions.

2.2.1 Setup

A sender S plays an asymmetric disclosure game with two receivers, an aligned receiver R_a, and a non-aligned receiver R_n. The sender (she) observes a private signal $\sigma \in \{\emptyset, h\}$ about the state of the world $\omega \in \Omega \equiv \{0, 1\}$ that is perfectly revealing: $Pr(\sigma = \emptyset|\omega = 0) = 1$ and $Pr(\sigma = h|\omega = 1) = 1$. We call the case that $\omega = 1$ ($\omega = 0$) the "high" ("low") state of the world. The choice of symbols for the two signal realizations, $\{\emptyset, h\}$, corresponds to the assumption that the sender can verifiably reveal her signal in one state of the world but not in the other: she can send a public message $m \in M \equiv \{\emptyset, h\}$ to both receivers

$$m \in \begin{cases} \{\emptyset\} & \text{if} \quad \sigma = \emptyset \\ \{\emptyset, h\} & \text{if} \quad \sigma = h. \end{cases}$$

That is, if the state is high, the sender can decide whether to pass her private signal on to the receivers ($m = h$) or to remain silent ($m = \emptyset$). If $\sigma = \emptyset$, she must remain silent $m = \emptyset$. A report of $m = h$ is, therefore, truthful and verifiable, whereas the empty message is not verifiable and may not be truthful. In this game, a disclosure strategy $\mu : \Omega \to \Delta(M)$, assigning to each state of the world ω a distribution over the messages $\mu(\omega)$, can simply be notated by a single number μ, the probability of not disclosing the high state if it occurs.

The two receivers have a common and correct prior $p = Pr(\omega = 1) \in (0, 1)$ and form posterior beliefs q_i upon seeing the sender's message.[4] Each receiver R_i with $i \in \{a, n\}$ then chooses an action $a_i \in [0, 1]$ to maximize

$$U_{R_i} = -(\omega - a_i)^2.$$

Given these payoffs, each receiver R_i aims for an accurate belief report and optimally chooses $a_i = q_i$. Without loss of generality, we therefore restrict attention to his posterior belief q_i.

The difference between the two receivers, R_a and R_n, is in the sender's utility from their actions. The sender wants R_a to know the state of the world; their interests are aligned. In contrast, the sender wants the non-aligned receiver R_n to think that the state of the world is

[4]As with the disclosure strategy, the posterior belief can be summarized by a single number – the belief in the "interesting" case $m = \emptyset$ – so we omit the argument when doing so causes no confusion.

low independent of the actual state of the world. The sender's payoffs are given by

$$U_S(q_a, q_n) \quad = \quad \alpha(1 - |\omega - q_a|) \quad + \quad (1-\alpha)(1-q_n)$$

where $\alpha \in [0,1]$ is a measure of the alignment of the sender and her entire audience: it governs the relative weight the sender attaches to each of the two receivers. Since the sender's message is public and visible to both receivers, her different persuasion goals toward the two receivers are in conflict with each other. However, a key property of the game's solution – to be tested in the experiment – is that the game does not allow for any differences in the two receivers' posterior beliefs q_a and q_n, as long as the players are rational and expect the same disclosure strategy μ.

In contrast to this rational prediction, we now consider the possibility of co-audience neglect. Co-audience neglect means that each receiver may over-weight his own importance in the game. A receiver with full co-audience neglect does not consider the presence of the other receiver. He is, therefore, blind to the other receiver's impact on the sender's incentives and, thus, on the informativeness of the message. Specifically, we propose the following solution concept to account for the hypothesized psychological bias. We call a tuple of disclosure strategies and posterior beliefs $(\mu, \tilde{\mu}_a, \tilde{\mu}_n, q_a, q_n)$ a Co-Audience Neglect Equilibrium (CANE) if

- R_i perceives S to use disclosure strategy $\tilde{\mu}_i$, in the sense that R_i's posterior belief after receiving message m, $q_i(m)$, is Bayes-rational given $\tilde{\mu}_i$
- R_a's perception of S's strategy, $\tilde{\mu}_a$, specifies that S chooses m to maximize

$$\tilde{U}_{S,a}(q_a) \quad = \quad 1 - |\omega - q_a(m)|$$

- R_n's perception of S's strategy, $\tilde{\mu}_n$, specifies that S chooses m to maximize

$$\tilde{U}_{S,n}(q_n) \quad = \quad 1 - q_n(m)$$

- S's actual strategy, μ, maximizes

$$U_S(q_a, q_n) \quad = \quad \alpha(1 - |\omega - q_a|) \quad + \quad (1-\alpha)(1-q_n)$$

  just as in the conventional (rational) equilibrium solution.

Thus we model co-audience neglect as a misperception of the sender's payoff function. While the receivers have non-rational beliefs, we assume that the sender has full knowledge of everyone else's payoffs and the receivers' misperceptions about her own payoff.[5]

2.2.2 Analysis

We first characterize the Bayesian Nash Equilibrium of the game, which will serve as a benchmark, and then turn to co-audience neglect. We call the sender's disclosure strategy truthful if she always reveals her private signal, i.e. $m = \sigma$, and deceitful if she sends $m = \emptyset$ regardless of her private signal. Simplifying the analysis slightly, we assume the sender to be truthful whenever she is indifferent. All proofs are relegated to Appendix B.1.

Proposition 1: In the unique Bayesian Nash Equilibrium of the game

- if $\alpha \geq \frac{1}{2}$, then the sender is truthful, $m^* = \sigma$, and receivers' equilibrium beliefs are accurate, $q_i^* = \omega$;
- if $\alpha < \frac{1}{2}$, then the sender is deceitful, $m^* = \emptyset$, and the receivers' equilibrium beliefs equal their prior, $q_i^* = p$.

Intuitively, if the aligned receiver is relatively more important to the sender, $\alpha \geq \frac{1}{2}$, then she is willing to bear the cost of letting the non-aligned receiver learn the truth; she, therefore, shares her information about the state of the world with the receivers. Conversely, if the non-aligned receiver is relatively more important to the sender, she is secretive and both receivers stick to their common prior.

Once again, note that receiver beliefs always coincide in Bayesian Nash Equilibrium. This ceases to be true if receivers neglect their co-audience. In that case, receivers will interpret the empty message $m = \emptyset$ differently. The aligned receiver ignores the sender's incentives to be secretive, expects the sender to always be truthful, and takes the empty message as perfect evidence of the low state of the world. The non-aligned receiver, in contrast, expects the sender to always be deceitful and deems the empty message to be uninformative about the state of the world.

Proposition 2: In the unique Co-audience Neglect Equilibrium of the game

[5] While the assumption of differential sophistication by role is strong, there is some evidence for asymmetric sophistication even on the individual level. Jin et al. (2021) present evidence in line with Forsythe et al. (1999) that laboratory subjects seem to behave strategically more sophisticated when assigned to the role of the sender, compared to the role of the receiver. Moreover, the vast psychological literature on egocentrism tends to support the assumption.

- if $\alpha \geq \frac{1-p}{2-p}$, then the sender is truthful ($m^{**} = \sigma$), R_a has equilibrium belief $q_a^{**} = \omega$, and R_n has equilibrium belief $q_n^{**} = p$ if $\omega = 0$ and $q_n^{**} = 1$ if $\omega = 1$;
- if $\alpha < \frac{1-p}{2-p}$, then the sender is deceitful ($m^{**} = \emptyset$), R_a has equilibrium belief $q_a^{**} = 0$, and R_n has equilibrium belief $q_n^{**} = p$.

When comparing the Bayesian Nash Equilibrium with the Co-audience Neglect Equilibrium, the following two corollaries immediately arise.[6]

Corollary 1: Co-audience neglect affects the sender's disclosure strategy.

In particular, the sender is truthful for a larger set of parameters (α, p) under co-audience neglect. Independent of co-audience neglect, the truthful strategy is more attractive to the sender when the average alignment between sender and receiver incentives α is high. Under co-audience neglect, a high common prior p makes the truthful strategy more attractive to the sender. For a graphical illustration of these comparative statics see Figure B.1 in Appendix B.

Corollary 2: Co-audience neglect polarizes beliefs. Upon seeing the empty message, the aligned receiver's posterior belief is strictly lower than the non-aligned receiver's posterior belief, $q_a(m = \emptyset) < q_n(m = \emptyset)$, for all values of α and p.

Thus, the model yields a clear prediction about the direction of belief polarization due to co-audience neglect. This directional difference in receivers' posterior beliefs upon the empty message is the key prediction for our experimental test of co-audience neglect. We can accommodate our framework for only a share $\gamma \in [0, 1]$ of receivers to be co-audience neglecting. For the clarity of the exposition, we outlined the two border cases $\gamma = 0$ (BNE) and $\gamma = 1$ (CANE) in the main text. The analysis of partial co-audience neglect can be found in Appendix B.1. Our main prediction that (on average) the aligned receiver has lower beliefs upon observing the empty message than the non-aligned receiver is obtained for all values of $\gamma > 0$.

[6] Note that sending the verifiable message, $m = h$, constitutes an out-of-equilibrium action to the non-aligned receiver. Since we model co-audience neglect as a misperception of the sender's payoffs rather than a misperception of the mapping from the sender's type (i.e. her private information) to the actions at her disposal (i.e. messages), the non-aligned receiver will know how to correctly interpret the verifiable message. Hence, beliefs upon seeing the verifiable message are degenerate and the same for all receivers regardless of co-audience neglect.

2.3 Experimental Design

At the core of the experiment is the interaction between three participants - one sender and two receivers - who are pursuing different goals, as randomly determined by their role. On an abstract level, this interaction has all the features of the disclosure game outlined in Section 2.2. This allows us to interpret the theoretical benchmarks in an empirical sense. We employ two treatments - one between-subject treatment and one within-subject treatment. The random role assignment, specifically the assignment to the role of the aligned versus non-aligned receiver, constitutes our main treatment. With this treatment, we test for co-audience neglect by comparing receivers' posterior beliefs by role. Remember that co-audience neglect, compared to the rational benchmark, implies different posterior beliefs when the sender's public message is received by people in different roles. In particular, aligned receivers are expected to hold lower posterior beliefs on average than non-aligned receivers. Any such directed difference in receivers' beliefs will be taken as evidence of co-audience neglect. In a second treatment, we vary the relative importance of the two receivers to the sender. While this treatment was mainly employed to make the sender's disclosure decision non-trivial, it lets us test for sender's rationality and everyone's understanding of the game. A deliberate feature of our experiment is that both receivers always have access to the exact same information set, which, from a standard perspective, precludes any difference in receivers' beliefs.

To ease participants' comprehension we use the metaphor of gold mining. The sender takes the character of a gold miner who finds gold (high state, $\omega = 1$) or does not find gold (low state, $\omega = 0$) with equal probability. The two receiver types are introduced as partner (aligned receiver) and bandit (non-aligned receiver). The message "gold was found" corresponds to the verifiable message, $m = h$, and the message "no gold was found" to the empty message, $m = \emptyset$.

Participants are randomly assigned to one of the three roles. Roles remain fixed throughout the session. Participants are given detailed and identical instructions: the true probability of the high state (common prior $p = \frac{1}{2}$), the decisions to be made by each player, their own incentives, and crucially, the incentives of players with different roles to their own. For eight rounds, participants are randomly matched into groups of three, where each group consists of one sender, one aligned receiver, and one non-aligned receiver. In each round, every group is randomly assigned to the high alignment condition or to the low alignment condition. In the high alignment condition, the sender assigns a higher weight to her incentives towards the aligned receiver than to her incentives towards the non-aligned receiver. The opposite is true for the low alignment condition. At the beginning of each round, everyone is informed about the alignment condition of their group. The sender then privately learns the state of the world. In the case of the low state of the world, the empty message is sent automatically to both receivers. Otherwise, the sender decides on sending either the verifiable message ($m = h$) or the empty message ($m = \emptyset$) to both of the receivers in her group. Simultaneously, receivers state how likely they think it is that the state of the world is high contingent on the sender sending the empty message. This belief, $q_i = P(\omega = 1|m = \emptyset)$, is our main outcome variable. Henceforth, if not specified otherwise, by belief, we refer to our main outcome variable. The sender then states her beliefs about those receiver beliefs (i.e. her beliefs on q_a, q_n). At the end of each round, participants are provided with rich feedback. The receivers learn the sender's message and the true state of the world. They do not learn the other receiver's belief. The sender learns both receivers' beliefs. Everyone learns their payoffs from that round.

Our incentivization ensures that the interaction among participants has all the strategic components of our disclosure game. The sender faces conflicting interests vis-a-vis the two receivers. She wants the aligned receiver to know the state of the world, whereas she wants the non-aligned receiver to have as low a belief as possible in the high state. Payoff functions are scaled such that each participant can earn a maximum of 400 lottery tickets that determine her chance of winning a prize of 8 euros. The sender's payoff depends directly on the receivers' posterior beliefs (q_a, q_n) and the state of the world

$$\text{Lottery Tickets}_\text{S} = 4\alpha(100 - |\omega * 100 - q_a|) + 4(1 - \alpha)(100 - q_n)$$

with $\alpha \in \{\frac{1}{4}, \frac{3}{4}\}$ and $q_i = P_i(\omega = 1|m = \emptyset) \in \{x \in \mathbb{N}_0 | x \leq 100\}$. Receivers use sliders to express, in percent, their perceived likelihood of a high state. Receivers do not have to state a belief contingent on the verifiable message, which directly implies the high state of the world (i.e. $q_i = P_i(\omega = 1|m = h) = 100$). The parameter α of the sender's payoff function determines the alignment condition, our second treatment. If a group in a specific round is randomly assigned to the high alignment condition, α is set to $\frac{3}{4}$ implying that the aligned receiver is relatively more important to the sender. In the low alignment condition, α is set to $\frac{1}{4}$ making the non-aligned receiver relatively more important to the sender. Hence, α governs the relative alignment between the sender's and receivers' incentives. Figure B.1 in Appendix B illustrates our parametrization of prior and alignment. Finally, senders can earn an additional bonus of 2 euros by correctly guessing receivers' beliefs. One of their two guesses is randomly chosen and if it is five percentage points or fewer away from the actual belief, the sender earns the bonus. In each session, one of the rounds is randomly determined to be payoff-relevant.

So that the hypothesized difference in receiver beliefs can be attributable to co-audience neglect, we took several measures to shut down alternative channels. If receivers, for instance, do not understand that the sender's message is public, beliefs would look exactly as they do under co-audience neglect. We therefore draw repeated attention to the fact that the message is public and ensured participants' knowledge of this with a comprehension question. Our online setting might have shrouded the presence of the other receiver (i.e. the co-audience). We tried to alleviate this concern by reminding every receiver of the presence of the other receiver at the beginning of each round when announcing their group's alignment condition. On top of that, there was a button on each page that, when clicked on, detailed the participant's role as well as every role's detailed incentives. The complete set of instructions can be found in Appendix B.

Procedural Details A total of 430 subjects participated in 9 online sessions in March and April 2021. Participants were students based in Munich, Germany, who were recruited via the online recruitment system ORSEE (Greiner, 2004). The experiment was implemented with oTree (Chen et al., 2016). Out of the 430 subjects who started the experiment, 70 did not pass the attention and comprehension checks within the allotted time. To ensure a smooth experience for the majority of participants, we had to employ relatively strict timers on each page. An additional 32 participants could not be matched into groups of three and were excluded. In total, 306 subjects completed the entire experiment. These 306 subjects, together with 22 subjects who were not able to complete all of the eight rounds, comprise our main sample ($n = 328$), see Table B.1 in Appendix B for a detailed breakdown. In this main sample, there are 108 senders, 110 aligned receivers, and 110 non-aligned receivers. A session lasted on average one hour, and participants who completed the entire experiment earned, on average, 14.10 euros. The experimental design, our hypotheses, and statistical analysis were pre-registered with AsPredicted (#58930). It is important to note that we pre-registered a sample of 240 participants to complete the entire study. This means that we ended up with more completions than pre-registered ($306 > 240$).[7] This oversampling was neither intended nor conditioned on preliminary results. All of the last five sessions were scheduled at the same time based on early lower completion rates. Strictly restricting our sample to the pre-registered sample size means excluding the last session. We decided to analyze all the data we collected but to be very transparent about how results change with the sample size (all data versus pre-registered sample size). In particular, every regression model featured in the main text will be presented with the pre-registered sample size in Appendix B. Any regression model that was not pre-registered will be flagged by an asterisk next to

[7]Including data from participants who did not complete the entire session but dropped out during the rounds in the analysis was pre-registered.

the dependent variable.

2.4 Results

Figure 2.1 shows receiver beliefs upon the empty message split up by role and by alignment condition. Independent of the role, beliefs are lower in the high alignment condition than in the low alignment condition, suggesting that receivers understand that the empty message is more informative in the high alignment condition. In the low alignment condition beliefs are, on average closer to the prior of 50%. Co-audience neglect predicts lower beliefs for the aligned receiver than for the non-aligned receiver in both alignment conditions. In most of the rounds, the beliefs of the aligned receivers are, on average, below those of the non-aligned receivers.

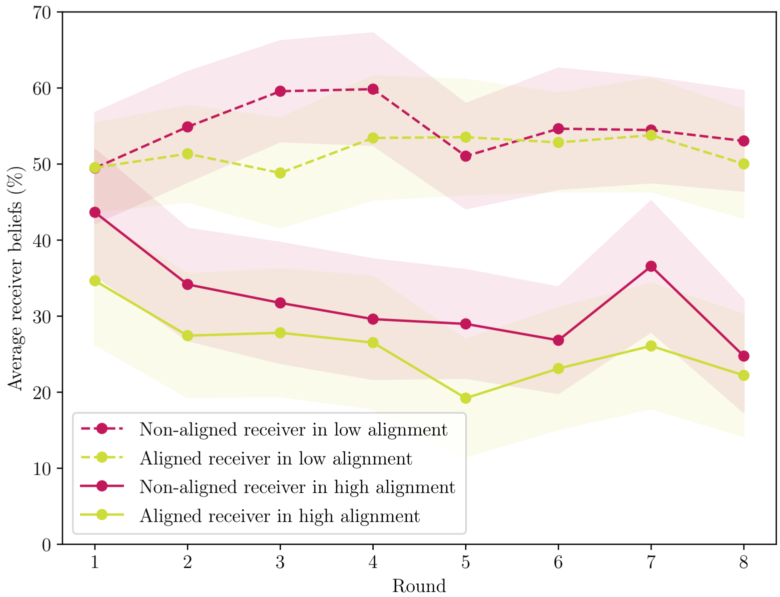


Figure 2.1: Average Receiver Beliefs Across Rounds

Notes: Average receiver beliefs upon the empty message. The shaded areas correspond to 95% confidence intervals.

Table 2.1 presents OLS regressions of receiver beliefs on our treatments. Aligned receivers believe the high state to be 4.5 percentage points less likely upon the empty message than non-aligned receivers (p-value = 0.06). This is evidence of co-audience neglect. This main treatment effect is not significant at any conventional level when considering only the pre-registered sample size (p-value = 0.13), see Table B.4 in Appendix B. Column (2) of Table

2.1 shows that beliefs are indeed significantly lower in the high alignment condition, though the difference in average beliefs is less than half of what the Bayesian Nash equilibrium would predict. The interaction of the high alignment condition and being the aligned receiver suggests that co-audience neglect is somewhat more severe in the high alignment condition, but this interaction is not significant. Surprisingly, column (3) shows that co-audience neglect is not less pronounced in later rounds after repeated feedback on senders' decisions. As column (4) shows, subjects do learn more readily along other dimensions. In particular, they are more responsive to the high alignment condition in the second half of the experiment. Column (5) then demonstrates that subjects naively take into account whether the state was high in the previous round. Another way of investigating learning, as opposed to only considering the number of rounds a subject has played, is to look at whether a receiver has observed a sender's decision yet. Since the sender only takes a disclosure decision if the state is high and the state is randomized at the group level, receivers had different exposure to feedback on senders' strategies at the same round. Note that whenever the state is high and the sender gets to decide on a message, her choice of message constitutes an out of Co-audience Neglect Equilibrium action to exactly one of the two receivers. If she decides to disclose (i.e. to send the verifiable message), this is an out-of-equilibrium action to the non-aligned receiver. If she decides not to disclose (i.e. to send the empty message), this is an out-of-equilibrium action to the aligned receiver. Using the idea of self-confirming equilibrium (Fudenberg and Levine, 1993) means that co-audience neglect should be sustained more easily as long as no sender decision has been observed. The variable "No disclosure experience (d)" measures whether a receiver has observed a disclosure decision. Countering the above intuition, column (6) shows that participants who have never received feedback on a sender's disclosure decision - which includes everyone in the first round - do not exhibit stronger co-audience neglect than those who did experience a disclosure decision.

As our balance table (Table B.2 in Appendix B) shows, treatment assignment was balanced on all observable characteristics except for net income. Our main treatment effect remains unchanged when controlling for this characteristic (see Table B.7 in Appendix B). More worrisome, the share of females assigned to the role of the non-aligned receiver was 10 percentage points higher than the share assigned to the role of the aligned receiver. While not statistically significant, this imbalance might have confounded our estimation of the main treatment effect. The main treatment effect, however, remains unchanged when controlling for the share of females. Our metaphor of gold mining may have framed the bandit as spiteful and might have led participants in that role to state high beliefs in order to "take money away" from the gold miner, the sender. When excluding all clearly irrational beliefs higher

Table 2.1: Receiver Beliefs

	(1) Belief	(2) Belief	(3) Belief	(4) Belief	(5) Belief*	(6) Belief*
Aligned receiver (d)	-4.537*	-2.939	-5.375*	-4.533*	-4.552*	-4.108
	(2.426)	(3.043)	(2.783)	(2.372)	(2.410)	(2.593)
High alignment (d)		-22.49***		-21.27***	-26.17***	
		(3.360)		(2.727)	(2.506)	
Aligned receiver x High alignment		-3.207				
		(4.798)				
Rounds 5 to 8 (d)			-4.099*	-0.391		
			(2.095)	(1.779)		
Aligned receiver x Rounds 5 to 8			1.706			
			(2.961)			
High alignment x Rounds 5 to 8				-5.736**		
				(2.852)		
High state at t-1 (d)					4.950**	
					(1.930)	
High state and disclosure at t-1 (d)					-3.447	
					(2.172)	
No disclosure experience (d)						2.923
						(2.564)
Aligned receiver x No disclosure exp.						-1.620
						(3.543)
Constant	43.40***	54.60***	45.41***	55.59***	54.48***	42.65***
	(1.752)	(2.208)	(2.046)	(2.132)	(2.084)	(1.842)
Observations	1676	1676	1676	1676	1456	1676
R^2	0.005	0.154	0.008	0.158	0.180	0.006

Notes: Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors are clustered at the participant level. One observation is one receiver belief in the high state upon the empty message, i.e. $q_i = P(\omega = 1|m = \emptyset)$. An asterisk next to an outcome variable indicates that the regression model was not pre-registered. Table B.4 in Appendix B displays this analysis for the pre-registered sample size.

than 50%, our treatment effect becomes smaller but more significant (see Table B.7 column 2 in Appendix B). Excluding those who clicked fastest through the instructions (fastest 10%) allows us to estimate the treatment effect more precisely (see Table B.7 column 6 in Appendix B).

Substantial across-subject variation suggests that the average treatment effect is driven by a subset of participants. We can classify a participant as a CAN type based on her belief stated in the first round. We call someone a CAN type if she stated a belief no more than 5 percentage points (10 percentage points) away from the Co-audience Neglect Equilibrium Prediction for her receiver type. By these definitions, 16% (20%) of our sample is classified as a CAN type. Column (1) of Table 2.2 shows that in the remaining rounds, co-audience is

Table 2.2: Heterogeneity in Receiver Beliefs

	(1) Belief rounds > 1*	(2) Belief rounds > 1*	(3) Belief*
Aligned receiver (d)	-2.332 (2.712)	-1.404 (2.772)	-11.88*** (4.431)
CAN Type (5pp deviation) (d)	-0.531 (3.913)		
Aligned receiver x CAN Type (5pp)	-15.12** (6.387)		
CAN Type (10pp deviation) (d)		-0.533 (3.610)	
Aligned receiver x CAN Type (10pp)		-15.98*** (5.627)	
Extreme political views			-2.924** (1.427)
Aligned receiver x Extreme political views			4.405** (2.015)
Constant	43.01*** (2.020)	43.03*** (2.085)	48.11*** (3.365)
Observations	1456	1456	1632
R^2	0.021	0.027	0.013

Notes: Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors are clustered at the participant level. One observation is one receiver belief in the high state upon the empty message (i.e. $q_i = P(\omega = 1|m = \emptyset)$). An asterisk next to an outcome variable indicates that the regression model was not pre-registered. Column (1) defines a CAN type as an aligned receiver that stated a belief $q_a \in [0, 5]$ in round 1 or a non-aligned receiver that stated a belief $q_n \in [45, 50]$ in round 1. Column (2) defines a CAN type as an aligned receiver that stated a belief $q_a \in [0, 10]$ in round 1 or a non-aligned receiver that stated a belief $q_n \in [40, 50]$ in round 1. Table B.5 in Appendix B displays this analysis for the pre-registered sample size.

only displayed by the subset of participants who were co-audience neglecting to begin with. This heterogeneity analysis is robust to a wider definition of a CAN type - see column (2).[8] Finally, column (3) shows that people with self-declared moderate political views exhibit more co-audience neglect. As political communication is generally geared towards moderates, this group may be less attuned to the idea that communication is often aimed at multiple audiences.

We find that senders react strongly to the alignment condition with 31% of disclosure decisions being truthful in the low alignment condition compared to 84% in the high alignment condition (see Table 2.3 columns (2) and (3) and Figure B.5 in Appendix B). Column (3) in Table 2.3 shows that senders are no more or less responsive to the alignment in the second half of the experiment. Another piece of evidence for senders' understanding of the

[8]This analysis is robust to various bandwidths used for the type classification, see Table B.8 in Appendix B.

Table 2.3: Sender Beliefs and Strategy

	(1) Sender belief	(2) Disclosure	(3) Disclosure
High alignment (d)	-3.505*	0.536***	0.555***
	(1.974)	(0.0501)	(0.0620)
W.r.t. aligned receiver (d)	-13.99***		
	(1.904)		
Rounds 5 to 8 (d)	-4.935		0.0221
	(3.401)		(0.0622)
W.r.t. aligned receiver x Rounds 5 to 8	0.662		
	(1.805)		
High alignment x Rounds 5 to 8			-0.0392
			(0.0750)
Constant	27.32***	0.306***	0.295***
	(3.371)	(0.0372)	(0.0448)
Observations	1662	418	418
R^2	0.071	0.294	0.294

Notes: Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors are clustered at the participant level. In column (1), one observation corresponds to one sender belief about one receiver belief, specifically her belief about $q_i = P(\omega = 1|m = \emptyset)$. In columns (2) and (3), one observation corresponds to one sender disclosure decision. Remember, a sender had to make a disclosure decision only when the state was high. Table B.6 in Appendix B displays this analysis for the pre-registered sample size.

game is that overall 78% of disclosure decisions coincide with the Bayesian Nash equilibrium prediction. Remember that parameters were chosen such that the sender's disclosure strategy is unaffected by co-audience neglect. Surprisingly, senders' disclosure decisions are somewhat more consistent with receivers' actual beliefs than the senders' own beliefs about receivers' beliefs (see Figure B.4 in Appendix B).

While, by design, we can not conclude from the senders' disclosure decisions whether senders expect co-audience neglect or not, sender beliefs about receiver beliefs are indicative of whether senders expect the receivers to be co-audience neglecting. Column (1) in Table 2.3 shows that senders indeed expect aligned receivers to hold significantly lower beliefs than non-aligned receivers (see Figure B.3 in Appendix B for a graphical representation of this result).

2.5 Conclusion

We have derived theoretical predictions of co-audience neglect and provided suggestive experimental evidence of the bias. The weak evidence should be assessed in light of the very strong empirical test employed, which arguably left little opportunity for the bias to be detectable. Outside of the asymmetric disclosure game studied here, co-audience neglect might be a source of loss of information in the public discourse, on top of social image concerns (Braghieri, 2021) or reputational concerns (Morris, 2001). To see this, consider a politician who is informed about the efficacy of an anti-discrimination policy by a social scientist. She might wrongly interpret the social scientist's public endorsement of the policy in question as unambiguous scientific backing of the policy, by failing to appreciate the public's impact on the scientist's verdict (i.e. its demand for politically correct views). This would lead to the loss of socially valuable information.[9] The ubiquity of echo chambers on social media, and news outlets narrowly catering to their readership, clearly limit the relevance of co-audience neglect for public communication, since a heterogeneous audience is a prerequisite for co-audience neglect to manifest itself.[10] Our design does not allow for any statements about the channels of co-audience neglect. It is likely that co-audience neglect works through selective attention to the sender's persuasion goals. Conceivably, listeners pay more attention to the sender's incentives pertaining to themselves, in the same way, that individuals pay more attention to asset prices of assets they own themselves (Hartzmark et al., 2021). It also remains an open question of how co-audience neglect interacts with the strategic complexity of the environment. Neglecting the co-audience could be a heuristic that individuals revert to when strategic communication becomes overly complex.

[9] This example is loosely based on the running example in Morris (2001).

[10] For economic discussions on echo chambers see for example Gentzkow and Shapiro (2011), Levy and Razin (2019), Oprea and Yuksel (2022).

Chapter 3

When Do Peers Influence Preventive Health Care Behavior? Evidence from Breast Cancer Screening

with Peter Redler

3.1 Introduction

Preventable diseases are a common cause of death in high-income countries (Mokdad et al., 2004; World Health Organisation, 2019). While preventive health care is commonly viewed as highly cost-effective, only a small share of individuals takes up preventive services (Borsky et al., 2018). The abundant supply of often-free preventive health care in high-income countries suggests that low take-up rates are primarily a demand-side issue.[1]

Numerous impediments to the demand for preventive health care have been discussed in the literature. Among them are a lack of or an inadequate delivery of information (Alsan et al., 2019; Torres et al., 2021; Alsan and Eichmeyer, 2023), pecuniary and non-pecuniary costs associated with participation (Banerjee et al., 2010; Campos-Mercade et al., 2021), flawed information processing (Loewenstein et al., 2013; Handel and Kolstad, 2015; Einav

[1] For example, in the United States, health insurance groups are obligated to provide coverage for any service recommended by the U.S. Preventive Services Task Force without any cost to the patient, regardless of the associated expenses.

et al., 2020), motivated avoidance of (health) information (Oster et al., 2013; Golman et al., 2017, 2022) and a lack of attention or awareness (Milkman et al., 2021; Dai et al., 2021). Some of these barriers are already addressed in preventive health care programs in high-income countries.[2] Others are difficult to address due to institutional constraints or welfare concerns.[3]

In search of new strategies to increase the take-up of preventive health care, we turn to the literature on peer effects and social dynamics in preventive health care behavior (Bouckaert et al., 2020; Francetic et al., 2022; Karing, 2023). We investigate the potential of social choice architecture, specifically the timing of check-up invitations and appointments, to activate these peer effects. We focus on breast cancer screening in Germany to answer the following two related research questions. Does inviting peers simultaneously to their breast cancer screening check-ups increase participation rates? When do peers influence an individual's decision to take up breast cancer screening?

The German breast cancer screening (BCS) program has many advantages to study these questions. First, BCS programs are an important instance of organized prevention efforts. Among women, breast cancer is the most prevalent cancer worldwide and responsible for a higher loss in disability-adjusted life years than any other cancer (Wild et al., 2020).[4] In an effort to decrease the morbidity and mortality of breast cancer, many high-income countries have established nationwide BCS programs. Due to EU-wide guidelines, the German BCS is highly comparable to other European BCS programs.[5] This extends the policy relevance of our results beyond the context of our study. Even though most of the aforementioned reasons for low take-up rates are addressed by the German BCS program, participation rates are stagnating at 50% (Deutsches Mammographie-Screening-Programm, 2022). As a consequence, there is a need to consider so far neglected aspects, such as the social choice architecture under which the individual makes her participation decision. Importantly, the German BCS program records participation decisions of all eligible women and demographic data that allows us to identify peers.[6] In Germany, all women aged 50-69 are offered free

[2]In our setting, for example, information is abundant and reliable, costs of participation are low, individual benefits are considered to be high (Katalinic et al., 2020), and measures are in place to minimize the psychological costs of receiving a positive diagnosis.

[3]Correcting the motivated avoidance of health information, for instance, is not unambiguously welfare-improving (Brunnermeier and Parker, 2005; Schwardmann, 2019). Often, an increase in awareness is achieved by information campaigns and reminders targeted at the individual, for example, through text messages (Milkman et al., 2021; Dai et al., 2021). Targeting individuals may, however, not always be feasible due to privacy concerns or institutional constraints.

[4]According to the Zentrum für Krebsregisterdaten (2021), 17% of all cancer-caused deaths among women in Germany were attributed to breast cancer in 2020. About one in eight women in Germany will develop breast cancer during their lives (Erdmann et al., 2021).

[5]Following the European Union's recommendations, as of 2020, 25 EU Member States have implemented population-based screening programs for breast cancer (European Commission, 2022).

[6]We will use address data and birthday data to construct peer networks.

biennial check-ups.[7] Eligible women receive a standardized letter inviting them to an appointment in roughly six weeks from the date they receive the letter.[8] We have access to data on 20,500 individuals residing in 19 villages in Germany.

Does inviting peers simultaneously to their breast cancer screening check-ups increase participation rates? We try to answer our first research question with a natural field experiment (Harrison and List, 2004) set within the German BCS program. Note that under the status quo, invitation letters and appointments are issued by birth date, so within a village, residents will receive their invitation letters at different times. By synchronizing invitation letters and appointments at the village level, our intervention seeks to turn the individual decision to participate in BCS into a more collective one. As a consequence, our intervention may increase take-up rates through several channels that have been documented in the literature on peer effects. In particular, our intervention may increase the awareness of BCS (Milkman et al., 2021; Dai et al., 2021) as well as the visibility of individual behavior (Bursztyn and Jensen, 2017), thereby activating signaling (Karing, 2023), influence (Esguerra et al., 2023) and conformity motives (Bernheim, 1994; Funk, 2010). Our experiment prevents us from distinguishing between these mechanisms, but it was designed to accommodate their potential involvement. For example, a woman may become more aware of BCS when an acquaintance mentions her own invitation or upcoming appointment. Our intervention ensures that this heightened awareness coincides with the point in time when the woman needs to make up her mind about BCS.

Specifically, our randomized controlled trial assigns 9 villages to the treatment group for which invitation letters are sent out simultaneously, and appointments are concentrated into the smallest possible time period. The remaining 10 villages are assigned to the control group, for which appointments are scattered across multiple weeks.

We document a precise null result with a minimal detectable effect size of 2.5 percentage points. Contrary to our hypothesis, synchronizing invitation letters and appointments does not increase participation rates. It then remains an open question whether there even are peer effects in our setting that could be leveraged.

When do peers influence an individual's decision to take up breast cancer screening? We try to answer our second research question with administrative data. The relative timing of peers' invitations and appointments may matter for an individual's likelihood to attend.

As argued above, peers who are invited at the same time as the individual may increase her awareness of breast cancer prevention. An individual could also be more likely to participate

[7]Women in this age group have a heightened risk of developing breast cancer, yet are still projected to derive gains from early detection and therapy.

[8]The proposed appointment does not have to be confirmed but can be rescheduled (see Section 3.2).

if more of her peers have their appointment on the same day as she does, for reasons such as being able to coordinate transport. Yet another possibility is that peers exert influence on the individual by showing behavior that the individual then wants to conform to (Bernheim, 1994; Funk, 2010): The share of peers who have their appointment prior to the individual's could positively influence the individual's likelihood of attending. Investigating these possibilities thus sheds light on the nature of peer effects in our setting. We identify peers based on spatial proximity and similarity in age using address data and birth dates. We then define several relative timing criteria (i.e., appointments on the same day, appointments within a week, etc.) and calculate the share of peers who fulfill the respective criterion. We can estimate the causal effect of the relative timing of peers' invitations and appointments on individual participation because of a property of the invitation algorithm that leads to exogenous variation in these peer shares.

We find that the probability that a woman will participate increases as the share of her peer group that has an appointment in the 7 days leading up to her appointment increases. This result is robust to several pre-registered definitions of peers. For example, an increase by one standard deviation in the share of peers residing within 500m of an individual who have their appointment scheduled in the 7 days leading up to her appointment increases her likelihood of participating by 1.7 percentage points. We do not find evidence for the effect of other relative timing criteria, such as being invited on the same day.

The results of our two empirical approaches are consistent. The share of peers invited on the same day does not increase an individual's likelihood of attending. Note that this is precisely the peer share that was targeted by our intervention. Instead, our peer share analysis suggests that some time needs to pass between peers' appointments for peers to be influential.

In addition, we document spatial autocorrelation in participation within villages. A proxy for socioeconomic status is positively associated with participation. One regional measure of social capital, the share of registered Christians, is also positively correlated with participation.

Until now, interventions to boost participation rates in BCS have primarily focused on modifying the invitation letter, yielding mixed outcomes (Goldzahl et al., 2018; Bertoni et al., 2020). Changing a letter is an extensively studied behavioral intervention (Robitaille et al., 2021; Allcott, 2011; Allcott and Rogers, 2014). It may, however, as in our setting, be infeasible.[9] In contrast, our equally cheap nudge changes the social choice architecture for the individual's decision (Benartzi et al., 2013).

[9]In Germany, national guidelines dictate the letter's exact wording.

Our analysis of peer effects makes several contributions. First, we take low-cost methods for approximating social networks that have been documented in other contexts (Drago et al., 2020; Beaman et al., 2021) and use them to deliver new insights into preventive health behavior. Second, we build on previous studies focusing on the peer influence exerted by spouses, co-workers and family members (Pruckner et al., 2020; Castro and Mang, 2022; Goldberg et al., 2022), and show that neighbors can also affect preventive health behavior.

Bouckaert et al. (2020) and Francetic et al. (2022) exploit discontinuities in the eligibility status of peers to identify peer effects. In their settings, peer effects are partly due to the preventive health care offer becoming more salient. We show that even conditional on all peers being eligible; peer effects exist - as long as the timing is right.

Recently, several studies on the welfare consequences of screening point to the detrimental effects of overdiagnosis (Einav et al., 2020; Kowalski, 2023). Our data preclude us from evaluating the welfare implications of BCS programs. More specifically, we are interested in identifying conditions conducive to high take-up rates of preventive health care offers.
In the next Section 3.2, we describe the German BCS program, the setting of our study. In Section 3.3, we describe the data we have access to. Section 3.4 contains the experimental design and the results of our intervention. In Section 3.5, we describe the empirical approach to our peer effect analysis and its results. Section 3.6 contains additional correlational findings on BCS participation. Section 3.7 concludes.

3.2 Background: Breast Cancer Screening in Germany

Based on European Union guidelines, the population-wide German breast cancer screening (BCS) program was established in 2005 and covers women aged 50 to 69 (Biesheuvel et al., 2011). Women in this age range who reside in Germany are invited every two years for a mammogram. This diagnostic procedure includes two X-rays of each breast from different angles. The images are independently assessed by two physicians, and any abnormal findings are referred for assessment by a specialist. If there is an abnormal finding and further testing is required, this is communicated within two weeks.

Invitation letters are sent out via mail, and the content of the letter is standardized (see Figure C.1 in Appendix C). A decision aid booklet which contains information on the procedure, on breast cancer in general, on possible outcomes, and on advantages and disadvantages of participation is enclosed with the invitation for first-time invitees.[10]

The letter contains a proposed appointment location, date, and time. This appointment

[10] Available at https://www.mammography-screening.de/download/downloads/broschueren/2019-08-13_G-BA_Entscheidungshilfe_Mammographie_EN_RZ_Web_2_2.pdf

slot is reserved for the woman, and confirming the appointment is optional. Invitation letters are typically sent out four to six weeks before the appointment. The recipient of the letter can request to reschedule the appointment. If a woman does not attend the screening, she receives one reminder with a proposed alternative appointment.

The mammograms are performed in screening units. While urban and suburban areas are typically served by outpatient locations, e.g., in hospitals or specialist practices, rural areas are routinely served by mobile mammography units (MMUs). These use equipment of similar quality to outpatient locations, but they only remain at each location for a few weeks. During this time, they serve as the screening location for all local women. The duration of a stay is determined by the capacity of the unit and the number of expected appointments for local women. Any area served by an MMU is typically visited every 24 months.

Goldzahl et al. (2018) and Carrieri and Wübker (2016) show that the screening programs in Europe have increased mammography rates in the relevant age groups. However, national participation rates vary widely (Wübker, 2014). In Germany, the overall uptake of screening within the program is around 50%, which is low in comparison to other EU countries and slowly decreasing over the past ten years (Deutsches Mammographie-Screening-Programm, 2022). Age-standardized breast cancer incidence and breast cancer mortality rates in Germany are slightly above the EU-28 country average (Dafni et al., 2019).

Mammograms are also performed outside of the BCS program, typically if symptoms such as pain or breast lumps are present or for preventive reasons, for example, given a family history of breast cancer. Therefore, the counterfactual to screening participation within the BCS program is not necessarily non-screening. Screening outside the BCS program is called opportunistic screening. The scope of opportunistic screening in Germany is not precisely known. The largest public German health insurance conglomerate reports that 8 to 12% of women in the target age range undergo a mammography exam outside the BCS program (Tillmanns et al., 2021). Nonetheless, screening within the BCS program might still be preferable to opportunistic screening due to higher diagnostic quality standards, i.e., better equipment and more experienced physicians.

3.3 Data

We use data from multiple sources. Our main dataset is administrative data from the German breast cancer screening (BCS) program. These data contain individual-level information on screening invitations, screening participation and participation history, age, and current address.

The sample area is a predominantly rural area in Germany (19 ZIP code areas in one federal state, approx. 200,000 inhabitants). Our sample comprises the universe of women aged 50 to 69 years in this area (n = 20,500). The program's regional structure corresponds to the administrative structure of Germany's ZIP codes: Women who reside in the same ZIP code are assigned to the same screening location.

For our natural field experiment (Section 3.4), the ZIP code areas correspond to treatment clusters. These areas are entirely non-urban: All towns in our sample have fewer than 50,000 inhabitants. Our data was collected over a period of several months (2022-2023) from one mobile mammography unit (MMU) that served 19 ZIP codes from four separate MMU locations. In the rest of this paper, we will refer to the 19 ZIP code areas as villages and to the four different MMU locations as sites 1-4.[11] Participation in BCS is our main outcome variable. We use previous participation and age as our main control variables. In an additional specification, we also control for the geographical distance between the individual's address and the MMU location and for weather, a school breaks dummy, and the current COVID-19 incidence.

For our *Peer Shares* approach (Section 3.5), we use individuals' birth dates and precise information on current addresses to create proxies for peer relationships. We combine these with the exact timing of the first proposed appointment to construct our explanatory variables.[12]

In addition, we ran two accompanying surveys to better understand the take-up decision process. At sites 1 and 2 of the trial, we distributed a paper-based survey at the screening site. For invitees at sites 3 and 4, we attached a QR code to the invitation letter that linked to an online survey similar to the paper-based survey. Both surveys were predominantly answered by women who participated in the screening.

We supplement these data with regional-level characteristics for heterogeneity analyses. We evaluate the role of social capital and social economic status, which we proxy with administrative data from the Federal Statistical Office. These data include voting data, data on religious affiliation, and local unemployment rates.

3.4 Intervention: Synchronizing Invitations

The purpose of our natural field experiment is to explore the potential of a simple, cheap, and scalable change to the invitation strategy to increase participation in breast cancer screening (BCS). In this section, we outline the design and implementation of our intervention and

[11] There is one town in our sample that has three ZIP code areas which we will treat as separate entities. Each one of the other 16 ZIP code areas corresponds to a geographically separate small area (a German *Gemeinde*).

[12] For more details, see Section 3.5.1.

present the main results and complementary survey evidence.

3.4.1 Design

We used random assignment to allocate villages to either the treatment or the control group. Each treated village was assigned the shortest possible time slot to accommodate all appointments of its residents. Appointments for women residing in untreated villages were scattered over a larger time interval, as determined by the status quo invitation algorithm. Since invitation letters are usually sent 6 weeks prior to the appointment, our intervention synchronizes the receipt of the invitation letters and the dates of appointments for the treatment group.[13] This alteration to the status quo invitation strategy intends to leverage peer effects. While we cannot discriminate among them, several channels through which our intervention could increase BCS participation seem plausible. First, our treatment may make individuals more aware of BCS. As an acquaintance mentions her own invitation or upcoming appointment, a woman may become more aware of her own invitation or upcoming appointment.[14] If, for instance, a woman was initially undecided or somewhat inattentive to the letter, such a conversation may nudge her toward participation. Second, our treatment may increase the visibility of individual behavior, which has been shown to encourage various target behaviors (Bursztyn and Jensen, 2017; Karing, 2023). If our intervention succeeds in making BCS a local topic of conversation, individuals' (intended) behavior should become more visible to others. An individual may be asked more often about her intention to participate. She may also realize through conversations that she is likely to be seen by peers at the check-up site. Third, simply observing more peers participating in BCS within a shorter time span of her own appointment may induce a woman to conform to this observed behavior (Bernheim, 1994; Funk, 2010). Alternatively, participating peers may serve as a timely reminder of a woman's own upcoming appointment. Lastly, our treatment may reduce participation costs by making neighborhood carpooling to the check-up site more feasible. Thus, we hypothesize that a woman is more likely to participate in BCS if more of her peers are invited at the same time as she is and have their appointments close in time to hers.[15]

Negative selection of screening participants on risk (Einav et al., 2020) makes it impor-

[13] One might be concerned about the lower level of privacy our treatment introduces. Women may not want to run into acquaintances or to be recognized by anyone at the screening. To evaluate this concern, we conducted a pre-intervention online survey with a representative sample of participants ($n = 170$). A vast majority of respondents (77%) indicated that they would not mind seeing an acquaintance at the check-up site. A smaller proportion of 12% indicated that such an encounter would indeed bother them, and 11% mentioned that they would enjoy meeting someone they know.

[14] BCS is a topic of conversation among invited women. In a representative online survey, 57% of women report that they talk with others about their decision to participate. Data from a second survey show that 88% of respondents believe that others talk openly about their participation in BCS.

[15] Our experimental strategy to test our main hypothesis requires a significant share of a woman's peers to reside in her village (see Section 3.3 for a discussion on using spatial proximity as a proxy for social ties).

tant to draw in previous non-participants. Investigating heterogeneous treatment effects by previous participation allows us to identify the effect of our intervention on this particularly policy-relevant subgroup of previous non-participants. The official invitation letter presents public health information in a formal manner, encouraging a woman to consider her personal costs, risks, and benefits associated with the check-up. Clearly, this invitation strategy fails to convince previous non-participants to take up BCS. These women may be more receptive to the informal communication channels that our intervention seeks to activate. Since our intervention only complements the invitation strategy currently in place and does not substitute any of its features, we do not expect any adverse effects on previous participants. The strong correlation between past participation and present participation suggests that any observed treatment effect on previous non-participants could also persist long-term.

Our sample comprises 20,500 women in 19 villages. During our intervention, the mobile mammography unit (MMU) served the local population at four different sites. These sites are typically chosen to be central and easily accessible, such as the parking lot of a shopping mall or a village square.

The impact of our treatment is limited by two factors. First, by law, all appointments must be offered within 22 to 26 months of a woman's previous appointment.[16] Second, given that an MMU only remains at each site for a couple of weeks or months at most, the appointments of women in our control group are also relatively close to each other.

Table 3.1: Experimental Design

Status quo - Appointments

<table>
<tr><th colspan="5">Week 1</th><th colspan="5">Week 2</th><th colspan="5">Week 3</th><th colspan="5">Week 4</th><th colspan="5">Week 5</th></tr>
<tr><td>M</td><td>Tu</td><td>W</td><td>Th</td><td>F</td><td>M</td><td>Tu</td><td>W</td><td>Th</td><td>F</td><td>M</td><td>Tu</td><td>W</td><td>Th</td><td>F</td><td>M</td><td>Tu</td><td>W</td><td>Th</td><td>F</td><td>M</td><td>Tu</td><td>W</td><td>Th</td><td>F</td></tr>
<tr><td colspan="25">All villages</td></tr>
</table>

Intervention - Appointments

<table>
<tr><th colspan="5">Week 1</th><th colspan="5">Week 2</th><th colspan="5">Week 3</th><th colspan="5">Week 4</th><th colspan="5">Week 5</th></tr>
<tr><td>M</td><td>Tu</td><td>W</td><td>Th</td><td>F</td><td>M</td><td>Tu</td><td>W</td><td>Th</td><td>F</td><td>M</td><td>Tu</td><td>W</td><td>Th</td><td>F</td><td>M</td><td>Tu</td><td>W</td><td>Th</td><td>F</td><td>M</td><td>Tu</td><td>W</td><td>Th</td><td>F</td></tr>
<tr><td colspan="5">Control villages</td><td colspan="4">Treated vill. 1</td><td colspan="7">Control villages</td><td colspan="3">Treated vill. 2</td><td colspan="6">Control villages</td></tr>
</table>

Notes: The status quo algorithm only considers dates of birth and the date of the last appointment. At a given site, the appointments of women from all surrounding villages are scheduled within a couple of weeks. The intervention reserves slots of a few days for treated villages (two in the example above). The capacity constraints of the MMU and the expected number of women attending their appointments determine the length of these slots. The remaining slots are filled with appointments of women residing in control villages according to the status quo algorithm.

Table 3.1 illustrates the scheduling of appointments under the intervention. Figure C.2 in Appendix C shows that this part of the manipulation, i.e., the concentration of appointments in treated villages, was successfully implemented across all four screening sites. The standard

[16] If a woman is invited for the first time, this range applies to her 50^{th} birthday.

deviation of the distribution of appointments within a village measures the degree to which appointments are dispersed over time. These are systematically smaller in treated villages than in untreated villages.[17] In both the treatment and the control group, around 15% of women rescheduled their appointments. Figure C.3 in Appendix C shows the final distribution of appointments after rescheduling. Again, the standard deviations of final appointment dates are systematically smaller in treated villages than in untreated villages.[18]

Usually, letters are sent out six weeks prior to the proposed appointment. Under the current practice, appointments are scattered over time, and accordingly, so is the letter's arrival. Our intention was to measure the combined effect of both receiving the letters *and* having the appointments close in time. As a result of an unintended deviation in the implementation process, all invitation letters for appointments at sites 3 and 4 (which correspond to the second half of the trial) were sent out simultaneously, regardless of treatment status.[19] While this was not intended and reduces our chance of precisely estimating our main treatment effect, it allows us to separately estimate (i) the effect of sending out the letters simultaneously and (ii) the effect of sending out the letters simultaneously and synchronizing the appointments.

3.4.2 Estimation

As pre-specified, we estimate the average treatment effect by the following equation

$$Y_{iv} = \beta_0 + \beta_1 T_v + \gamma prevpart_i + \delta age_i + \epsilon_{iv}. \tag{3.1}$$

Y_{iv} represents a binary participation variable, and $prevpart_i$ is a dummy that equals 1 if a woman participated in the screening previously (at least once during 2018-2021). We also control for age. Previous participation is highly predictive of our outcome and thus increases the power of our design.

We cluster our standard errors at the village level, our level of randomization. Because there are at most 19 clusters in our setting, we use the wild-cluster bootstrap proposed by Cameron et al. (2008) and use Rademacher-weights, as suggested by Canay et al. (2021). We report bootstrapped p-values in addition to our estimates.[20]

The following estimation interacts previous participation with the treatment dummy to detect heterogeneous treatment effects by previous participation.

[17] The average standard deviation of initially proposed appointment dates weighted by village size within the treated (untreated) villages is 5.24 days (14.71 days).

[18] The average standard deviation of final appointment dates in the treatment (control) group is 15.13 days (21.01 days).

[19] Table C.1 in Appendix C details how the intervention differs between the first and the second half of the trial.

[20] We follow the approach of Alan et al. (2023) who face a similar econometric setting.

$$Y_{iv} = \beta_0 + \beta_1 T_v + \beta_2 T_v \times prevpart_i + \gamma prevpart_i + \delta age_i + \epsilon_{iv}. \tag{3.2}$$

3.4.3 Results

On average, women who previously participated and women who did not previously participate in the BCS are unaffected by our intervention. Figure 3.1 shows participation rates by treatment status and invitation type. Among the cohort that was invited to the BCS for the first time, we do, however, observe a significantly lower participation rate in the treatment group (6.3% difference, p-value = 0.05). We have no information on past participation for women who received their first invitation to the screening during our intervention. Per our pre-registration and due to substantial variation in participation across cohorts within villages, we exclude first-time invitees from the primary analysis. Controlling for previous attendance drastically reduces the variance in observed participation.

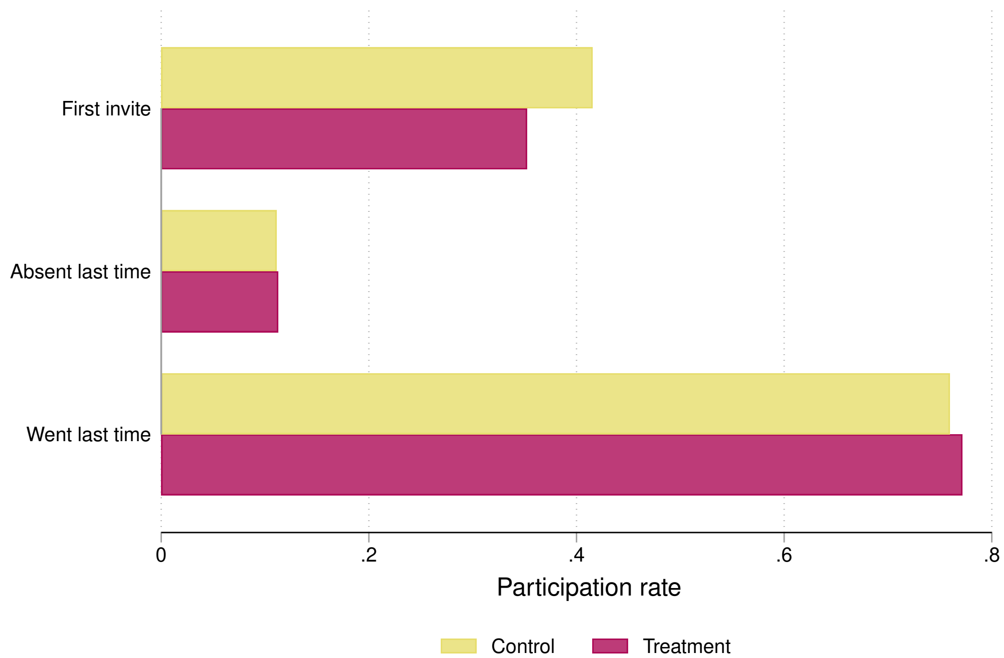


Figure 3.1: Participation Rate by Previous Attendance

Notes: Previous attendance, *Absent last time* v. *Went last time*, refers to participation in 2018-2021 (i.e., a woman did or did not attend during this period). During the trial, 11% of women were invited for the first time, *First invite*. Previous attendance has been known to be a strong predictor of participation (Deutsches Mammographie-Screening-Programm, 2022).

Table 3.2 presents the experimental results. Throughout all specifications, previous participation is strongly associated with participation. We separately estimate (i) the effect of simultaneous appointments (columns (1) - (3)), (ii) the effect of simultaneous letters (columns (4) - (6)), and (iii) the combined effect of simultaneous appointments and letters (columns (7) - (9)) on participation. Within each of the three groups of columns, the leftmost column presents results from Estimation 3.1. The middle column includes a dummy for previous participation as in Estimation 3.2. In the rightmost column, control variables are added.[21]

Simultaneous appointments (Table 3.2 columns (1) - (3)): There is no overall effect of bunching appointments on participation. The participation rate of residents of villages who are invited within as narrow a time frame as possible is not different from villages whose residents' appointments are scattered over time. Throughout all columns, the point estimates for the interaction effects are positive but not significantly distinguishable from zero. We thus do not find evidence for a heterogeneous treatment effect by previous attendance.

Simultaneous letters (Table 3.2 columns (4) - (6)): Due to the accidental treatment arm discussed in Section 3.4.1, we now isolate the effect of receiving the invitation letters simultaneously within a village. To this end, we are comparing the control villages of the first two sites to the control villages of the last two sites. Appointments were not bunched for any of these villages. The control villages of the last two sites, however, received invitations at the same time. Thus, they act as treated villages for this analysis. Sending out letters simultaneously without bunching appointments naturally generates variability in the time interval between the receipt of the letter and the appointment date. The variable *lead time* captures this gap in time. As one might expect, the relationship between lead time and participation is negative. We do not find evidence that sending out the invitation letters at the same time increases participation.

Simultaneous letters and appointments (Table 3.2 columns (7) - (9)): We document no effect of our originally planned intervention on participation. Again, this effect does not mask heterogeneous treatment effects by previous attendance.

Table C.3 in Appendix C restricts the analysis to first-time invitees. In line with the observed raw difference in participation rates by treatment (see Figure 3.1), we find significant and negative treatment effects for this new cohort. We argue for caution in interpreting these results for several reasons. First, because we cannot control for prior participation, statistical power becomes a more prominent concern, leading to a presumably higher intra-cluster correlation (ICC). Second, as we observe only one cohort per village, the sample size within some villages is very small ($n<50$ for 6 villages). This might further elevate the ICC, given that participation rates are correlated within cohorts.

[21] Reassuringly, treatment assignment is balanced across all control variables, except for precipitation. Table C.2 in Appendix C shows that control villages have a larger population and are located closer to their screening site, but these differences are not statistically significant.

Table 3.2: Main Results - Dependent Variable: Participation

	Simultaneous Appointments			Simultaneous Letters			Sim. Appointments + Letters		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Treated	0.006	0.000	-0.002	-0.007	-0.007	-0.019	0.001	-0.006	-0.012
	(0.010)	(0.006)	(0.010)	(0.007)	(0.016)	(0.013)	(0.015)	(0.013)	(0.019)
Went last time * Treated		0.011	0.007		0.000	0.007		0.012	0.012
		(0.022)	(0.021)		(0.029)	(0.026)		(0.027)	(0.027)
Lead time				0.000	0.000	-0.000	0.000	0.000	-0.000
				(0.000)	(0.000)	(0.001)	(0.000)	(0.000)	(0.000)
Went last time	0.658^{***}	0.654^{***}	0.659^{***}	0.653^{***}	0.653^{***}	0.655^{***}	0.660^{***}	0.653^{***}	0.654^{***}
	(0.012)	(0.015)	(0.014)	(0.015)	(0.021)	(0.022)	(0.014)	(0.021)	(0.022)
Age	-0.002^{**}	-0.002^{**}	-0.002^{**}	-0.002^{**}	-0.002^{**}	-0.002^{*}	-0.001	-0.001	-0.001
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Precipitation, in cm			0.006			0.003			0.004
			(0.006)			(0.009)			(0.005)
Avg. temperature (C)			0.000			-0.001			-0.001
			(0.000)			(0.002)			(0.001)
Distance to MMU (km)			0.001			0.001			0.001
			(0.001)			(0.001)			(0.001)
School break			-0.011^{*}			-0.013^{*}			-0.013
			(0.006)			(0.006)			(0.012)
Covid 7d inc.			-0.007^{**}			-0.008^{**}			-0.006^{*}
			(0.003)			(0.003)			(0.003)
Constant	0.223^{***}	0.225^{***}	0.247^{***}	0.215^{***}	0.216^{***}	0.282^{**}	0.180^{**}	0.184^{**}	0.241^{***}
	(0.043)	(0.044)	(0.049)	(0.050)	(0.043)	(0.102)	(0.063)	(0.064)	(0.049)
Observations	18336	18336	18095	11042	11042	10816	11362	11362	11284
R^2	0.429	0.429	0.432	0.422	0.422	0.426	0.432	0.432	0.433
Clusters	19	19	19	10	10	10	15	15	15
WC Bootstrap p-value	0.541	0.825	0.950	0.332	0.715	0.475	0.962	0.907	0.880

Notes: Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors are clustered at the village level. Wild-Cluster Bootstrap p-values for (joint) significance of treatment effects are reported. Columns (1) – (3) compare all control villages (*Treated*= 0) with all treatment villages (*Treated*= 1). Columns (4) - (6) compare control villages of the first half of the trial (*Treated*= 0) to control villages of the second half of the trial (*Treated*= 1). Columns (7) - (9) compare control villages of the first half of the trial (*Treated*= 0) with all treatment villages (*Treated*= 1). See Table C.1 in Appendix C for a breakdown of how the intervention differs between the first and the second half of the trial. Results for first invites are reported separately in Appendix C.

The null result of the intervention needs to be interpreted in the context of two limitations. First, there is a statistical power constraint, as the number of clusters is small, and the size of the clusters varies significantly. The confidence interval of the main treatment effect estimates indicates that we can only rule out effect sizes larger than approximately 2.5 percentage points. Since the implementation costs of synchronizing letters and appointments within villages are negligible, even a true effect size of 1 or 2 percentage points would be relevant and might warrant changes to the invitation strategy.[22]

The second main limitation is the specific setting of our study, namely the rural area served by mobile mammography units. Collaborating with MMUs was partly driven by logistical considerations. There are at least two reasons why our intervention is likely to yield stronger results in a less rural setting. First, an initial pilot study carried out prior to the national roll-out of breast cancer screening reports a substantially higher participation rate at mobile screening units than at non-mobile screening units (Kolip and Wurche, 2005).[23] Assuming that it becomes harder to draw in marginal non-participants at higher participation rates, increasing participation rates at MMUs is more difficult than at stationary units. Second, as Figures C.2 and C.3 document, appointments at MMUs fall into relatively short time spans, even absent our intervention. Consequently, our intervention can be interpreted as very light touch, further reinforcing conditions that we expected to be conducive to high take-up rates.

3.4.4 Survey evidence

We ran two surveys alongside the intervention to complement our understanding of the decision environment and to explore mechanisms such as awareness, visibility, communication, and coordination. During the first half of the trial, we distributed an on-site paper-based survey among BCS participants (Survey 1). During the second half of the trial, we attached a QR code for an online survey (Survey 2) to the invitation letter.[24] In both cases, the treatment status of respondents is inferred by the self-reported residence village. Table 3.3 details sample sizes and response rates.

At least two assumptions have to hold to be able to attribute any differences in a survey item by treatment status to our intervention. First, there are no baseline differences by treatment status with respect to the survey item. Second, there is no differential selection by treatment in taking the survey. Both assumptions are likely to be violated in our setting.[25]

[22] Synchronizing appointments requires a small change in the appointments-scheduling algorithm.

[23] Unfortunately, the program's yearly assessment reports do not break down participation rates by mobile units and stationary units, but anecdotally, participation rates are still thought to be higher at mobile units.

[24] Since the medical on-site personnel reported that administering the on-site survey (Survey 1) interrupted their workflow, we were unable to extend the paper-based survey to the second half of the trial and employed the online survey (Survey 2) instead.

[25] Due to the low number of clusters (i.e., villages) and strong variation of survey participation rates over

Table 3.3: Surveys - Overview

	N	Sample	Response rates			Mode	BCS participation
			Overall	Treatment	Control		
Survey 1	946	BCS participants	0.27	0.4	0.18	on-site	1
Survey 2	629	BCS invitees	0.1	0.1	0.11	online	0.9

Notes: For Survey 2, BCS participation is self-reported (intended) participation. Prior to the intervention, we ran an exploratory online survey ($n = 170$) with a separate and representative sample. All survey materials are in Appendix C.3.

Thus, the following results are to be interpreted as consistent with, but not necessarily confirming, our approach.

Our survey evidence suggests that our intervention increased the perceived knowledge of peers' participation behavior and the perceived openness towards conversations around BCS. It also suggests that there is scope for unambiguously welfare-improving interventions that increase take-up rates.

Survey 1

Survey 1 elicited an individual's migration history, education, previous participation in breast cancer check-ups - within and outside of the BCS program, perceptions about acquaintances' participation in BCS, mode of transport to the appointment, and whether or not she would like to attend her appointments more regularly or not.

Table C.4 in Appendix C documents that only migration status differs by treatment status, which likely reflects baseline differences. The positive signs of the coefficients in columns (1) and (2) indicate that, as intended, our intervention somewhat increased knowledge about how many acquaintances are participating in the BCS. Column (1) shows that in the treatment group, 3 percentage points more women report knowing what share of their acquaintances are participating in the BCS (p-value = 0.18). Interestingly, respondents perceive that roughly 75% of their acquaintances participate in the BCS compared to an overall participation rate of 47%. Since all Survey 1 respondents are BCS participants, this suggests a strong correlation in peers' participation behavior, which is consistent with results presented in Section 3.6.

Our intervention may have lowered the costs of reaching the check-up site by making carpooling more feasible. We document, however, that baseline rates of carpooling are very low at 3% and are not affected by our treatment. There is a significant and negative correlation (p-value < 0.01) between having participated in a BCS check-up and a check-up outside of

time (see Figure C.4 in Appendix C), both assumptions are likely to be violated despite randomized treatment assignment.

the BCS (i.e., opportunistic screening), suggesting that BCS reaches women who otherwise would not proactively engage in breast cancer prevention.

Survey 2

The vast majority of Survey 2 responses (83%) were collected prior to the appointment, and almost 90% of respondents were either planning to go to their check-up or reported having gone already. In addition to all questions included in Survey 1, we asked respondents how they perceived others' willingness to talk about breast cancer prevention check-ups.
First, we confirm two findings from Survey 1. Only 0.4% of survey takers are carpooling to reach the MMU.[26] We again document strong substitutability between BCS and opportunistic screening (p-value $< .01$). Notably, 22% of Survey 2 respondents say that they would like to attend the biennial check-ups more regularly. This implies that the policy goal of increasing participation rates is aligned with individual interests and that an intervention to target these individuals could be welfare-increasing.

Since most respondents take the survey right after receiving the letter, i.e. before they could be affected by our intervention, we now differentiate between pre-appointment and post-appointment responses. The results are displayed in Appendix C.1 (Table C.5 - C.7). In line with our exploratory survey, breast cancer prevention is not a stigmatized topic: 88% of respondents believe that others talk openly about their participation in BCS. This belief is held more often in the treatment group (2.8% difference, p-value = 0.01) and is more pronounced, although not significant when restricting the sample to post-appointment responses (see Table C.6 in Appendix C). We interpret this increased openness towards conversations about breast cancer prevention check-ups as a sign that, in our setting, peers can be used to disseminate public health information.

Consistent with both the exploratory survey and Survey 1, only 38% of respondents state that they have no idea how many of their peers are participating in the BCS. When restricting the analysis to post-appointment responses, we find that our intervention significantly increased the share of women reporting knowledge of their acquaintances' participation behavior (p-value = 0.06, see Table C.7 in Appendix C). This suggests that our intervention has given participants a better understanding of their peers' behavior.

[26] A vast majority of 70% reports going alone by car.

3.5 Peer Effects on Participation

Our pre-registered *peer shares* setup aims to evaluate whether and when peer influence matters for an individual's decision to participate in BCS. We first describe our setup in detail and then present the results.

3.5.1 Setup

In our setup, we consider all women in one village as potential peers to each other. We use two dimensions as proxies for peer relationships to classify peers: Geographic proximity and age proximity. We then test whether the relative timing of proposed invitation dates between those peers affects participation.

Relative timing might matter because the dates of appointments are closely linked to the dates of letter receipt, as the letters are usually sent a fixed period before the appointments. This creates opportunities for women to influence each other's decisions through conversations when they receive the letter, just before the appointment, or at any point in between. The decision to participate remains flexible, as appointments neither need to be confirmed nor actively canceled. This results in ongoing potential for peer effects.

We employ two alternative approximations of peer relationships that generate unweighted and undirected peer networks. Our first approximation is based on address data. The underlying assumption is that geographic proximity is related to the likelihood of peer interactions. In a similar setting to ours, rural Austria, Drago et al. (2020) document a high communication intensity among neighbors that declines monotonically with geographical distance. Marmaros and Sacerdote (2006) find that geographic proximity is a more important determinant of peer interactions than interests or family background, further substantiating our assumption. Whenever two women live within a given cutoff distance from each other, we consider them as linked.[27] We vary the cutoff distance.

Our second approximation relies on the relative age of peers. The underlying rationale is a homophily argument (McPherson et al., 2001), which suggests that women of similar ages are more likely to interact with each other, possibly due to shared age-based experiences such as attending school together. As a narrower third approach, we combine the two previous dimensions and define peers as women who live within a specified distance and have birth dates that are within a specified range of months. We end up with several approximations of peer networks, all of which were pre-registered. Combining these with pre-registered relative timing criteria, we can analyze along which dimensions peer effects matter for the decision

[27] We calculate the haversine distance, i.e., the shortest distance over the earth's surface between two coordinates.

Table 3.4: Criteria to Define Peers

Peer criterion	**Dimensions**
Geographical distance d	$d < \bar{d}$ meters, $\bar{d} \in \{50, 200, 500, 1000\}$
Time between birth dates k	$k < \bar{k}$ years, $\bar{k} \in \{0.5, 1, 2\}$
d and k	Combinations of criteria above (4 x 3 = 12)

Notes: The peers of a woman are defined as all women who live fewer than $\bar{d}$ meters away, have an age difference of less than $\bar{k}$ years, or a combination of both criteria (i.e. women who live less than $\bar{d}$ meters away and have an age difference of less than $\bar{k}$ years). These criteria were pre-registered.

to get a mammogram. Having several approximations of peer networks allows us to evaluate the robustness of any effect.

For each invited individual i, we calculate the number of peers n who fulfill each peer criterion. Table 3.4 lists the specific cutoffs used as peer criteria for our peer effects models.

Given a set of n peers, we then analyze the role of the relative timing of the proposed screening dates. We do this based on the assumption that the relative timing of the proposed date is related to peer interactions. Intuitively, if two women receive the letter on the same day and have an identical proposed screening date, this could increase the likelihood of an interaction among them that relates to breast cancer screening. Table 3.5 shows the relative timing criteria used.

We pre-registered the timing criteria to evaluate the role of the timing of peer interactions in the participation decision. For example, if carpooling was a main mechanism of peer interactions, the criterion *Same day* might be of key importance. If peers who had their appointment in the days leading up to the individual's appointment served as role models, the criterion *1 to 7 days before* might matter more. We also include the two criteria *More than 7 days before* and *More than 7 days after* that we do not expect to strengthen peer effects as placebo criteria.

We then construct peer shares for each individual i as the share of n women who fulfill a relative timing criterion. Figure 3.2 illustrates how a peer share using peers as determined by geographical proximity is determined. All women living within distance $\bar{d}$ of woman i are counted, here $n = 10$. Then, the number of peers who fulfill the relative timing criterion are counted, here $x = 4$. This results in a peer share $share_i = 0.4$. If the relative timing criterion was *Same day*, this would imply that 40% of individual i's peers have their appointment on the exact same day as individual i.

Peer shares are a function of the timing of the proposed screening dates. They vary based on the geographical location of the woman, the age of the woman and the proposed appointment. Table C.8 in Appendix C shows how the calculated peer shares vary. For

Table 3.5: Relative Timing Criteria

Relative timing	**Intuition**
Same day	Similar timing of letter, exact same date of appointment
Within two days	Similar timing of letter and appointment
Before	Peer had appointment before i (excludes same day)
1 to 7 days before	Peer had appointment closely before i
More than 7 days before	Gap between appointments, peer's appointment earlier (Placebo)
More than 7 days after	Gap between appointments, peer's appointment later (Placebo)

Notes: To construct a peer share for a woman i, we calculate the share of a set of peers who fulfill a binary relative timing criterion c. The sets of peers vary by specification as defined by Table 3.4. The relative timing refers to the peer's proposed initial appointment relative to the proposed initial appointment of woman i. The intervals for the criteria *Before*, *More than 7 days before* and *More than 7 days after* are only bounded on one side.

smaller sets of defined peers (e.g. $d < 50m$), the values mechanically diverge more.

For exogeneity of the set of our key explanatory variable $share_i$, the proposed appointment time needs to be random conditional on previous participation and distance to the screening location. The source of this randomness is the invitation algorithm. The algorithm proposes appointment dates. It uses information on previous participation of the invited women so that utilization levels of the screening site do not deviate strongly from planned utilization levels. This is accomplished by splitting invitees into three groups: First invites, those who participated last time, and those who did not participate last time. A mix of invitees from these groups is then assigned to each time slot. Crucially, within our groups, the algorithm sorts primarily by birth date or by the date of previous invitation. The order of appointments for first invites within a village is thus determined by the order of the birth dates. Because participation is plausibly unrelated to the birth order within a village, appointment timing is exogenous within villages for first invites. Since the previously proposed screening date again matters for the appointment in subsequent years, the source of exogeneity is passed on within cohorts across screening years. Since appointment timing directly affects whether a relative timing criterion is fulfilled (see Table 3.5), the peer shares are in turn determined orthogonally to potential participation.

Due to an inadvertent change in the invitation algorithm because of our experimental intervention (see Section 3.4), this exogeneity argument does not hold for treated villages. In order to schedule appointments for the treated group, the original invitation algorithm, which sorts primarily by age, was replaced with an algorithm that primarily sorts by address. The resulting peer shares are thus not orthogonal to potential participation.[28] We, therefore,

[28]Sorting by address likely results in endogeneity of the share variable. We show in Section 3.6 how a proxy for SES is related to participation.

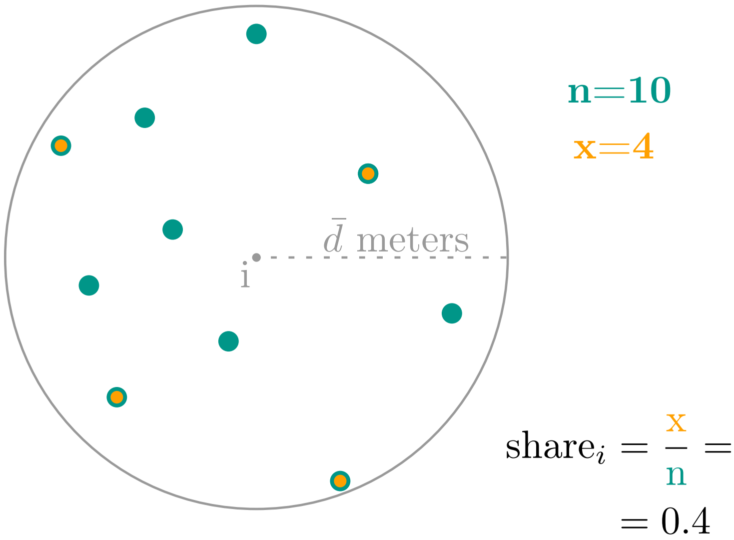


Figure 3.2: Geographical Peer Shares Calculation

Notes: All n women residing within (haversine) distance of $\bar{d}$ meters of woman i are considered her peers. Among these peers, all x women whose appointments fulfill relative timing criterion c with respect to woman i's appointment are counted. This count, x, is divided by the number of peers, n, in order to calculate woman i's peer share, $share_i \in [0, 1]$.

exclude all treated villages from the peer share analysis.

Given our setup, we estimate the following linear probability model for each version of the constructed peer shares. Y_{iv} represents screening participation that equals 1 if a woman participates in the screening at any point during our study period,

$$Y_{iv} = \beta_0 + \beta_1 share_i + \gamma prevpart_i + \pi_v + X\beta + \epsilon_i. \tag{3.3}$$

Our coefficient of interest is β_1, which can be interpreted as the change in the participation rate in % given an increase in $share_i$ of 1%. We control for previous participation as well as age, distance from residence to the screening location, and the absolute number of peers n. Due to village fixed effects, any estimates can be understood as effects relative to the average woman in a municipality.

3.5.2 Results

We first present results for estimating Equation 3.3. We flexibly vary the construction of $share_i$ as previously discussed. The variable varies across the peer dimensions geographic distance $\bar{d}$ and age difference $\bar{k}$ and the relative timing of the invitations (see Section 3.5.1).

Figure 3.3 shows the estimated β_1 coefficients for different versions of Equation 3.3 where the peer criterion is geographic distance. Each set of a point estimate and its 95% confidence interval represents a separate regression. Detailed regression results are presented in Tables C.9 - C.14 in Appendix C.

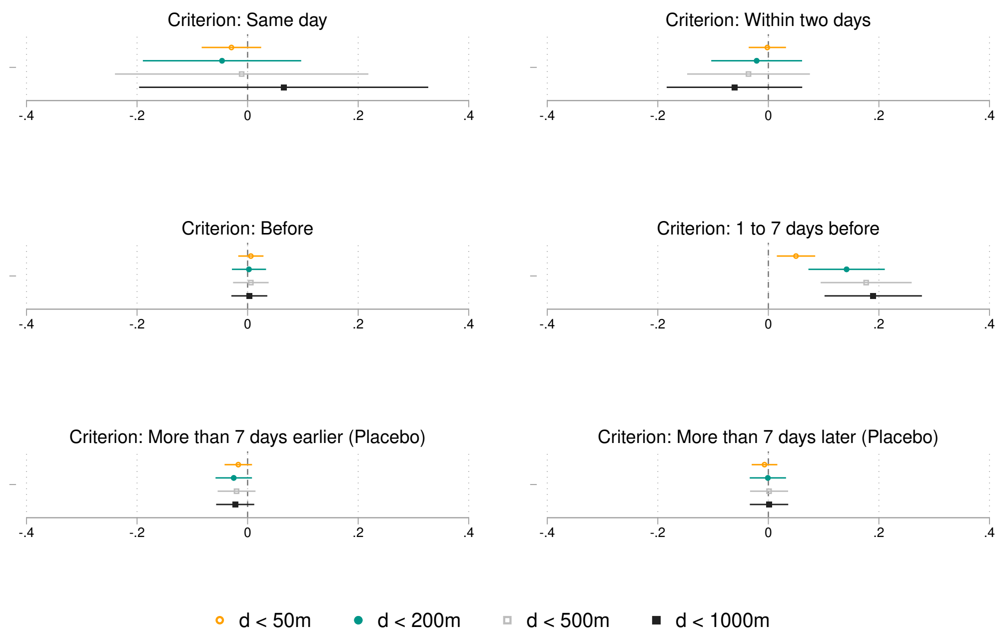


Figure 3.3: Peer Shares by Geographic Distance

Notes: This figure presents resulting β_1 from estimations of Equation 3.3. Each bar and its standard errors represent one regression that includes village fixed effects and controls for previous participation, the size of the peer group, distance to the mammography unit, and age. The header of each panel indicates the relative timing criterion (see Table 3.5). Each bar per panel represents a different peer criterion, here by distance in meters (see Table 3.4). The full regression results are presented in Tables C.9 - C.14 in Appendix C.

Most coefficients are statistically indistinguishable from zero. Notably, the results for the criterion *Same day* suggest that identical timing of appointments does not play a role. This aligns with our survey result that carpooling is uncommon (see Section 3.4.4).

The panel for *1 to 7 days before* shows significant positive estimates of the coefficient of interest. This reflects that when more peers of an individual within a specified geographic distance are invited to the screening the week before the individual, her own participation probability increases. These effects are relevant in size. For example, the point estimate for $d < 500m$ suggests that an increase in this peer share by one standard deviation (0.093, see Table C.8 in Appendix C) leads to an estimated increased participation probability of 1.7 percentage points. Coefficients for $d < 50m$ are closest to zero. This is not surprising given that the number of peers within such a close distance is often small. Given the construction of the explanatory variable $share_i$, extreme values such as 0 or 1 are comparably more likely. These extreme values, in combination with a smaller number of peers, likely bias estimates

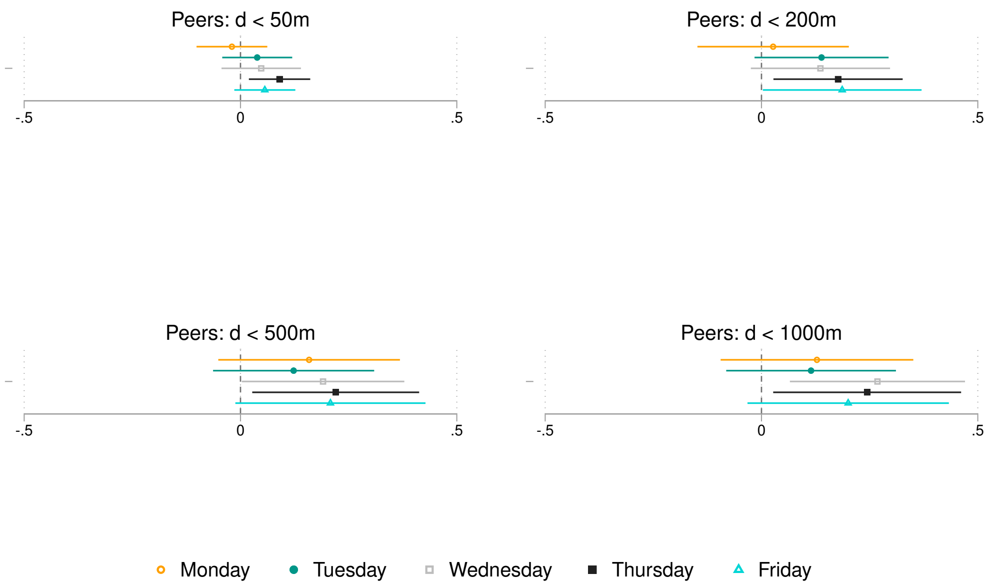


Figure 3.4: Coefficients for Relative Timing Criterion *1 to 7 days before* by Day of the Week

Notes: This figure presents β_1 estimates of Equation 3.3 separately by day of the week of the invited woman's initial appointment. Each bar and its standard errors represent one regression that includes village fixed effects and controls for previous participation, the size of the peer group, distance to the mammography unit, and age. The header of each panel states the peer criterion (see Table 3.4). Each bar per panel represents a different day of the week. The relative timing criterion used here is *1 to 7 days before*.

towards zero.[29]

To explore heterogeneity within the results for the criterion *1 to 7 days before*, we split our sample by weekday of the proposed appointments in Figure 3.4. While confidence bands increase, the relevant β_1 point estimates are lower when the initial appointment is earlier in the week (Monday or Tuesday). The results suggest that appointments later in the week are more susceptible to peer influence. A mechanism consistent with smaller peer effects on Monday and Tuesday is the lack of opportunity to rearrange one's week. For example, if a woman is influenced by a peer on the weekend - the time when typically more leisure is available, which results in peer interactions - she might only be able to make time to attend her appointment if it is later in the week.

Figure 3.5 is analogous to Figure 3.3, but the strategy to proxy peers differs. Now, women of similar age are considered her peers. Specifically, the distance between the birth dates of woman i and her potential peers is considered. As above, the share of peers that fulfill a relative timing criterion with respect to their screening appointment determines the values of

[29] Another aspect that biases these estimates toward zero is that we control for the number of peers linearly.

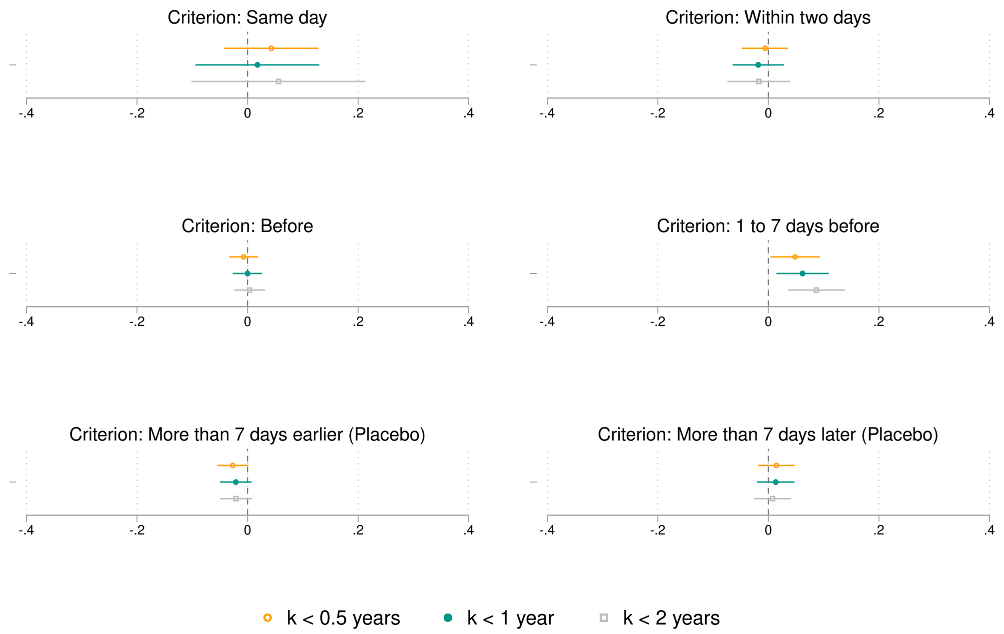


Figure 3.5: Peer Shares by Age Distance

Notes: This figure presents β_1 from estimations of Equation 3.3 which can be interpreted as the change in participation in % given an increase in $share_i$ by 1%. Each point estimate and its 95% confidence bands represent one regression that includes village-fixed effects and controls for previous participation, the size of the peer group, distance to the mammography unit, and age. The header of each panel indicates the relative timing criterion (see Table 3.5). Each bar per panel represents a different peer criterion, here, by difference in age (see Table 3.4). The full regression results are presented in Tables C.15 - C.20 in Appendix C.

$share_i$.

Here, a similar overall picture emerges. Again, the estimates for the timing criteria *1 to 7 days before* are positive. While the estimates are smaller in absolute terms, they are still significantly different from zero. Defining the age difference cutoff differently does not play a large role for the estimates of any relative timing criterion.

We combine the peer criteria geographical distance and age distance in Figures C.5 - C.7 in Appendix C. Here, only individuals who live both within distance $\bar{d}$ and are close in age are considered peers. The relative timing criteria remain unchanged. This results in less statistical power. However, the point estimates do not change much compared to the previous results: Again, we observe positive estimates for the relative timing criterion *1 to 7 days before* that are mostly statistically significantly different from zero.

Overall, these results support the hypothesis that peers serve as role models and influence an individual's decision to attend when their appointments are just before the individual's appointment. Salience might also play a role: If a woman observes that a peer participates in the screening, it might serve as a reminder for her own appointment. Because the estimates for the peer criterion *Before* are close to 0, it does not seem to matter only whether a peer's appointment was before the individual's own appointment: Proximity matters.

3.6 Spatial Correlation and Social Determinants

3.6.1 Spatial correlation

This section presents additional findings that we did not pre-register. First, we implement a variation of the *peer shares* approach from Section 3.5. We estimate the association between *participation rates* of peers and own participation conditional on a set of covariates. Again, geographical distance and time between birth dates are the criteria that we use flexibly to proxy relevant peers (see Table 3.4). The criterion used to calculate the *participation peer share* is now different: It is simply the share of peers that participated in the screening during our study period and thus not related to the relative timing of invitation dates.

This exercise should not be interpreted as causal as there is no plausibly exogenous variation in the participation rates of peers. It also represents a case of the reflection problem (Manski, 1993). We estimate equations of the following linear probability model

$$Y_{iv} = \beta_0 + \beta_1 share_i + \eta_i + \pi_v + X\beta + \epsilon_i. \tag{3.4}$$

η_i represents invitation-type effects as our sample now also includes individuals who were invited for the first time. We further control for village fixed effects, distance to the site, age, and the number of peers. Our coefficient of interest, β_1 coefficients, can be interpreted as follows: An increase in the participation rate of the peers by 1% is associated with a change in their own participation probability of β_1%. Due to the village fixed effects, our estimates are always relative to the average woman in the village.

The estimates reported in Table 3.6 are positive and significant for all distances other than 1000m (for which the point estimate is also positive). This represents a significant positive spatial autocorrelation within villages.

In a similar vein, we use address data to construct spatial grids of all villages. We construct square grids with sides 100 meters in length and calculate the participation rates in a given grid cell. For each grid cell, we define all adjacent grid cells to the north, south, west, and east as neighboring cells. We then calculate the Moran's I statistic that tests for spatial autocorrelation between participation rates against the null hypothesis of random spatial

Table 3.6: Participation Peer Shares - Peers by Location

	(1) 50m	(2) 200m	(3) 500m	(4) 1000m
Share	0.047***	0.11***	0.20***	0.074
	(0.0095)	(0.025)	(0.042)	(0.058)
Age	-0.0033***	-0.0032***	-0.0032***	-0.0031***
	(0.00058)	(0.00056)	(0.00056)	(0.00056)
n within distance	-0.0041***	-0.00020	-0.000083**	-0.000032**
	(0.00088)	(0.00014)	(0.000035)	(0.000015)
Constant	0.56***	0.51***	0.49***	0.54***
	(0.034)	(0.035)	(0.039)	(0.044)
Village + Inv. Type FE	Yes	Yes	Yes	Yes
Observations	18970	20315	20388	20421

Notes: Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. This table presents results from estimating Equation 3.4 with variations of the peer criterion distance (in meters).

autocorrelation.[30]

Table 3.7: Spatial Autocorrelation in Participation: Moran's I

Minimum n	Grid cells	Moran's I	p-value
1	5547	0.005	0.389
2	4084	0.039	0.032
3	3031	0.052	0.022
4	2220	0.069	0.018
5	1595	0.056	0.085
6	1131	0.105	0.022
7	807	0.074	0.128
8	560	0.057	0.248
9	391	0.041	0.351
10	280	0.046	0.358

Notes: This table presents the results for the calculation of Moran's I for the participation rate across adjacent pairs of grid cells. The grid cells are 100 meters x 100 meters. In each row, all grid cells with less than the minimum n of women per grid cell are disregarded, thus gradually reducing the number of grid cells and accordingly the number of pairs. The reported p-values are presented against an assumed random distribution of participation rates.

Table 3.7 shows the resulting p-values. These p-values can be interpreted as the probability of observing a spatial pattern as extreme as the one found in the data if the null hypothesis is true. We also impose different conditions on the minimum number of women within a grid cell to be regarded in the calculation of Moran's I. The null hypothesis of no spatial autocorrelation at a very close geographical level can thus be rejected.

[30] We cannot present the heat maps of participation rates for data protection reasons.

Next, we test for age-based autocorrelation in participation. We estimate versions of Equation 3.4 where the peer group is now defined as all individuals whose age differs by a maximum of $\bar{k}$ years. The coefficients can be interpreted as the predicted difference of the individual's participation probability compared to the average woman in a village conditional on controls.

Results are reported in Table 3.8. The point estimates are indistinguishable from zero, indicating no age-based autocorrelation within villages. They also do not change much depending on the specified maximum age distance.

Table 3.8: Participation Peer Shares - Peers by Age

	(1) 0.5 years	(2) 1 year	(3) 2 years
Share	-0.013	-0.014	-0.030
	(0.042)	(0.059)	(0.084)
Age	-0.0035***	-0.0038***	-0.0039***
	(0.00067)	(0.00070)	(0.00073)
n within age relation	-0.00017	-0.00015*	-0.000089**
	(0.00016)	(0.000081)	(0.000039)
Constant	0.61***	0.65***	0.67***
	(0.062)	(0.063)	(0.061)
Village + Inv. Type FE	Yes	Yes	Yes
Observations	20432	20432	20432

Notes: Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. This table presents results from estimating Equation 3.4 with variations of the peer criterion age difference (in years).

Finally, we combine the criteria geographical distance $\bar{d}$ and age distance k. Now, the relevant peers in the denominator of the $share_i$ variable from Equation 3.4 are individuals who live within distance $\bar{d}$ and whose birth date is less than a specified number of $\bar{k}$ months away from the individual's birth date. Table 3.9 presents the β_1 estimates for different combinations of $\bar{d}$ and $\bar{k}$. The estimates are consistently positive and significant, stressing the strength of the spatial correlation. The effect sizes are comparable to those in Table 3.6.

3.6.2 Socioeconomic status and social capital

Finally, we evaluate associations of participation rates with additional variables that relate to socioeconomic status and social capital. First, we include voting data. Specifically, we use the turnout in the previous national election as a measure of political participation. Second, we use the local unemployment rate as a proxy for regional socioeconomic deprivation. We also calculate the share of the adult population that is registered as catholic or protestant

Table 3.9: Participation Peer Shares - Peers by Distance and Age

	(1) 50m	(2) 200m	(3) 500m	(4) 1000m
0.5 years	0.0260**	0.023***	0.027**	0.023
	(0.013)	(0.008)	(0.013)	(0.017)
1 year	0.020**	0.026***	0.072***	0.083***
	(0.010)	(0.010)	(0.017)	(0.023)
2 years	0.025***	0.032***	0.091***	0.096***
	(0.009)	(0.013)	(0.023)	(0.031)

Notes: Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. This table presents results from estimating Equation 3.4 with variations of the peer criterion across two dimensions, i.e., by both geographical distance (column) and age difference (row). Only β_1 estimates are reported. Control variables and fixed effects are the same as in Tables 3.6 and 3.8.

Christians, which we interpret as a proxy for social capital (Strømsnes, 2008; Traunmüller, 2011).

At the individual level, we also use the number of other invited women who live within 50 meters as a measure of the population density of the individual's immediate neighborhood. As larger lot sizes are negatively associated with this proxy - especially in a non-urban setting - this is a proxy for socioeconomic status. Other research has used lot sizes and property values as a proxy for SES (Juhn et al., 2011; Ware, 2019). Lower SES women are also presumably more likely to live in the same apartment building or block as others, while higher SES women tend to live in larger properties.[31]

Table 3.10 shows that women who live in more densely populated areas are less likely to participate, suggesting a positive relationship between socioeconomic status and participation at the individual level. For the regional variables, only the share of registered Christians is positively associated with participation. This is in line with Salmon et al. (2022) reporting a positive association between church attendance and screening in the United States. Women in regions with more of this type of social capital are seemingly more likely to participate in the screening.

[31]A SES gradient in mammography participation has been documented in other settings, e.g., by Khan et al. (2021) in Australia or Lemke et al. (2015) in Germany. Smith et al. (2019) provide an overview.

Table 3.10: Heterogeneity in Participation Rates

	(1)	(2)	(3)	(4)
SES proxy (density)	-0.390***		-0.340**	-0.387**
	(0.094)		(0.102)	(0.103)
Unemployment (per cent)		0.010	0.010	0.002
		(0.014)	(0.014)	(0.015)
Election turnout		0.001	-0.013	0.054
		(0.108)	(0.110)	(0.094)
Share of registered Christians		0.181*	0.122	0.176*
		(0.065)	(0.075)	(0.066)
Constant	0.410***	0.239	0.309*	0.277*
	(0.018)	(0.115)	(0.121)	(0.112)
Invitation-type FE	Yes	Yes	Yes	Yes
Additional controls	No	No	No	Yes
Observations	20719	20719	20719	20432
R-squared	0.381	0.381	0.381	0.384

Notes: Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. This table presents results from regressing individual participation on a set of individual and village-level variables. The SES proxy (density) is defined as the number of other invited women who live within 50 meters. Unemployment, election turnout, and the share of registered Christians is available at the village level. Additional individual-level control variables include weather, distance to the site, school breaks, and COVID-19 7-day incidence.

3.7 Discussion

In this paper, we provided causal evidence that the relative timing of peers' preventive check-ups matters for individual participation: A higher share of an individual's peers who have an appointment in the days leading up to her own appointment increases her likelihood of participating in BCS. The results from our intervention indicate that synchronizing invitation letters and appointments does not increase participation rates. Since our intervention maximized the expected share of peers who have their appointment on the same day, these results do not contradict each other. One possible explanation is as follows.

The sequence of social signals that a woman may be exposed to from the time she is invited until her appointment can further reconcile our results. If there are conversations about BCS, receiving the letter at the same time as the entire village exposes an individual to mixed signals: With participation rates at 50%, in expectation half of the conversation partners will share an intention to participate and the other half will share an intention not to participate. The situation differs starkly once the individual's appointment is imminent. Since the act of participation is more salient than the act of non-participation, the individual will be mainly exposed to positive signals on the value of BCS.[32] Evidence from behavioral economics suggests that people underreact to empty signals - in our setting, the act of non-participation (Tversky and Kahneman, 1973; Enke, 2020; Jin et al., 2021). Thus, close to her appointment, a woman may overestimate participation in BCS and be inclined to conform to the perceived norm of participating (Bernheim, 1994; Funk, 2010). Participating peers may also remind her of her own upcoming appointment. In principle, this dynamic could have unfolded as a result of our intervention. Our peer share analysis, however, suggests that some time needs to pass between peers' appointments and the individual's appointment for peers to be influential. This longer time spell may be required for conversations to take place and for individuals to be able to react to peer influence.

We have identified the timing of check-ups as an unexplored dimension that can be manipulated even within a large-scale, tightly regulated preventive health care program. While we still consider the optimal design of screening invitation campaigns to be an open question, we hope that our identification of the scheduling as an important and mutable feature of these programs may prove useful to future research and policy design.

[32] Arguably, a peer is more likely to share that she just went to or is going to her appointment rather than the fact that she could have gone to an appointment that she was invited to six weeks ago and decided against.

Appendix to Chapter 1

A.1 Mathematical Appendix

A.1.1 Proofs

Equilibrium under Coarse Moral Language

In equilibrium, all agents of type $v \in [\hat{v_r}, v_r)$ take action r whereas all other agents of type $v \notin [\hat{v_r}, v_r)$ take their natural action. The cutoff type $\hat{v_r}$ is pinned down by the fixed point equation

$$U(r, \hat{v_r}) - U(a(\hat{v_r}), \hat{v_r}) = \hat{v_r}e(r) - c(r) - (\hat{v_r}e(a(\hat{v_r})) - c(a(\hat{v_r}))) + \gamma\Delta(\hat{v_r}) = 0, \tag{A.1}$$

where $\Delta(\hat{v_r}) = E(v|v \geq \hat{v_r}) - E(v|v < \hat{v_r})$. I first show that this action profile constitutes an equilibrium and then proceed by showing that the equilibrium exists and is unique.

Claim 1: The above action profile constitutes a Perfect Bayesian Equilibrium.

Proof of Claim 1: Let's call $ve(r) - c(r) - (ve(a(v)) - c(a(v)))$ the direct utility gain of choosing action r over action $a(v)$ for agent v. The derivative of this direct utility gain with respect to v is given by

$$\begin{aligned} &\frac{\partial ve(r) - c(r) - (ve(a(v)) - c(a(v)))}{\partial v} \\ &= e(r) - e(a(v)) - v\frac{\partial e(a(v))}{\partial v}\frac{\partial a(v)}{\partial a} + \frac{\partial c(a(v))}{\partial v}\frac{\partial a(v)}{\partial a} \\ &= e(r) - e(a(v)) + \frac{\partial a(v)}{\partial a}(\frac{\partial c(a(v))}{\partial a} - v\frac{\partial e(a(v))}{\partial a}). \end{aligned} \tag{A.2}$$

By $a(v)$ being the natural action, $v\frac{\partial e(a(v))}{\partial a} - \frac{\partial c(a(v))}{\partial a} = 0$. Therefore,

$$\frac{\partial ve(r) - c(r) - (ve(a(v)) - c(a(v)))}{\partial v} = e(r) - e(a(v)) \begin{cases} > 0 & \iff v < v_r \\ = 0 & \iff v = v_r \\ < 0 & \iff v > v_r. \end{cases} \tag{A.3}$$

To prove that the action profile constitutes an equilibrium, let's consider all types separately and show that there is no profitable deviation available to them.

- Take $v < \hat{v_r}$. Then, as Equation A.3 shows, the direct utility gain of choosing action r over action $a(v)$ is increasing in v. This implies that $U(r, v) - U(a(v), v) < 0$, i.e., the agent prefers her natural action $a(v)$ over action r. Any other action would make her worse off.

- Take agent $\hat{v_r}$. Since, $U(r, \hat{v_r}) - U(a(\hat{v_r}), \hat{v_r}) = 0$, she is indifferent between choosing her natural action $a(\hat{v_r})$ and taking action r. By assumption, she chooses action r. Any other action (other than $a(\hat{v_r})$) would make her worse off.

- Take $v \in (\hat{v_r}, v_r)$. Still, as Equation A.3 shows, the direct utility gain of choosing action r over action $a(v)$ is increasing in v. This implies that $U(r, v) - U(a(v), v) > 0$, i.e., the agent prefers r over her natural action $a(v)$. Any other action would make her worse off.

- Take $v > \hat{v_r}$. Deviating from her natural action $a(v)$ would make her worse off.

No restrictions on out-of-equilibrium beliefs are needed to sustain this Perfect Bayesian Equilibrium since I only consider conventions that use both categories in equilibrium.

□

Claim 2: The above equilibrium exists and is unique.

I impose two assumptions.

1. $$v_{min}e(r) - c(r) - (v_{min}e(a(v_{min})) - c(a(v_{min}))) + \gamma\Delta(v_{min}) < 0 \tag{A.4}$$

2. $U(r, v) - U(a(v), v)$ is strictly monotone in v, i.e. $e(r) - e(a(v)) + \gamma\frac{\partial \Delta(v)}{\partial v} > 0$. This requires γ not be too large (in the case that $\frac{\partial \Delta(v)}{\partial v} < 0$).

Proof of Claim 2: Proving existence is equivalent to proving that there exists a v such that Equation A.1 holds. Note that the following expression is true as long as $\gamma > 0$

$$\underbrace{v_r e(r) - c(r) - (v_r e(a(v_r)) - c(a(v_r)))}_{=0} + \gamma\Delta(v_r) > 0. \tag{A.5}$$

By Assumption A.4 and Equation $A.5$, $U(r, v_{min}) - U(a(v_{min}, v_{min}), v_{min}) < 0$ and $U(r, v_r) - U(a(v_r), v_r) > 0$. By the assumed monotonicity of $U(r, v) - U(a(v), v)$ in v, there exists a $\hat{v_r} \in (v_{min}, v_r)$ such that $U(r, \hat{v_r}) - U(a(\hat{v_r}), \hat{v_r}) = 0$. The strict monotonicity makes $\hat{v_r}$ the unique solution of the equilibrium condition (Equation A.1).

□

Effect of a Language Designer's Objective Function

The language designer's objective function is given by

$$o(f(v), c(a), l(a), r) = \int_{v_{min}}^{\hat{v_r}} l(a(v))f(v)dv + \int_{\hat{v_r}}^{v_r} l(r)f(v)dv + \int_{v_r}^{v_{max}} l(a(v))f(v)dv. \tag{A.6}$$

Claim 3: Let $m(a, r^*)$ be the optimal convention that maximizes $o(f(v), c(a), l(a), r)$ under linear $l(a)$. If $\tilde{l}(a)$ is concave, the optimal convention must not define the negative moral category more broadly than the optimal convention under $l(a)$ (i.e., $\tilde{r}^* \leq r^*$).

Proof of Claim 3: By $m(a, r^*)$ be the optimal convention under $l(a)$ the following holds

$$\begin{gathered} o(f(v), c(a), l(a), \tilde{r}^*) \leq o(f(v), c(a), l(a), r^*) \\ \int_{\hat{v}_{\tilde{r}^*}}^{v_{\tilde{r}^*}} (l(\tilde{r}^*) - l(a(v)))f(v)dv \leq \int_{\hat{v}_{r^*}}^{v_{r^*}} (l(r^*) - l(a(v)))f(v)dv \\ \frac{\int \int_{\hat{v}_{\tilde{r}^*}}^{v_{\tilde{r}^*}} (l'(\tilde{r}^*) - l'(a(v)))f(v)dvdv}{\int \int_{\hat{v}_{r^*}}^{v_{r^*}} (l'(r^*) - l'(a(v)))f(v)dvdv} \leq 1 \end{gathered} \tag{A.7}$$

Assume by contradiction that $\tilde{r}^* > r^*$. Under the linear welfare function $l(a)$, Equation A.7 holds for $\tilde{r}^* > r^*$. If the linear welfare function $l(a)$ is replaced with the concave $\tilde{l}(a)$ in Equation A.7, by concavity of $\tilde{l}(a)$, the expression holds with strict inequality,

$$\frac{\int \int_{\hat{v}_{\tilde{r}^*}}^{v_{\tilde{r}^*}} (\tilde{l}'(\tilde{r}^*) - \tilde{l}'(a(v)))f(v)dvdv}{\int \int_{\hat{v}_{r^*}}^{v_{r^*}} (\tilde{l}'(r^*) - \tilde{l}'(a(v)))f(v)dvdv} < 1. \tag{A.8}$$

Thus, under a concave value function $\tilde{l}(a)$, the language designer would strictly prefer $m(a, r^*)$ to $m(a, \tilde{r}^*)$ which is a contradiction to $m(a, \tilde{r}^*)$ being the optimal convention under $\tilde{l}(a)$.

Claim 4: Let $m(a, r^*)$ be the optimal convention that maximizes $o(f(v), c(a), l(a), r)$ under linear $l(a)$. If $\tilde{l}(a)$ is convex, the optimal convention must not define the negative moral category more narrowly than the optimal convention under $l(a)$ (i.e., $\tilde{r}^* \geq r^*$).

Proof of Claim 4: The proof of Claim 4 is analogous to the proof of Claim 3.

□

A.1.2 Example of Ambiguous Total Effect

Let's assume a simple, linear externality function, $e(a) = a$, and a quadratic cost function $c(a) = \frac{1}{2}a^2$. If $\gamma = \frac{2}{5}$, an agent v's utility function is given by

$$U(a, v) = va - \frac{1}{2}a^2 + \frac{2}{5}E[v|m(a, r)]. \tag{A.9}$$

I assume that types are normally distributed, with $\mu = 3$ and $\sigma^2 = 1$, i.e. $v \sim f(v) = \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}(v-3)^2}$. I evaluate the effect of changing from a narrow convention (r_n) to a broad convention (r_b, $r_b > r_n$) on aggregate moral behavior.

1. If r increases from $r_n = 2$ to $r_b = 4$, the diluting effect dominates the purifying effect, and the total effect on actions as in Equation 1.2 is negative.

2. If r increases from $r_n = 4$ to $r_b = 5$, the purifying effect dominates the diluting effect, and the total effect on actions as in Equation 1.2 is positive.

This shows that the total effect of broadening a negative moral category is ambiguous. Figure A.1 illustrates the ambiguous effect using simulated data.

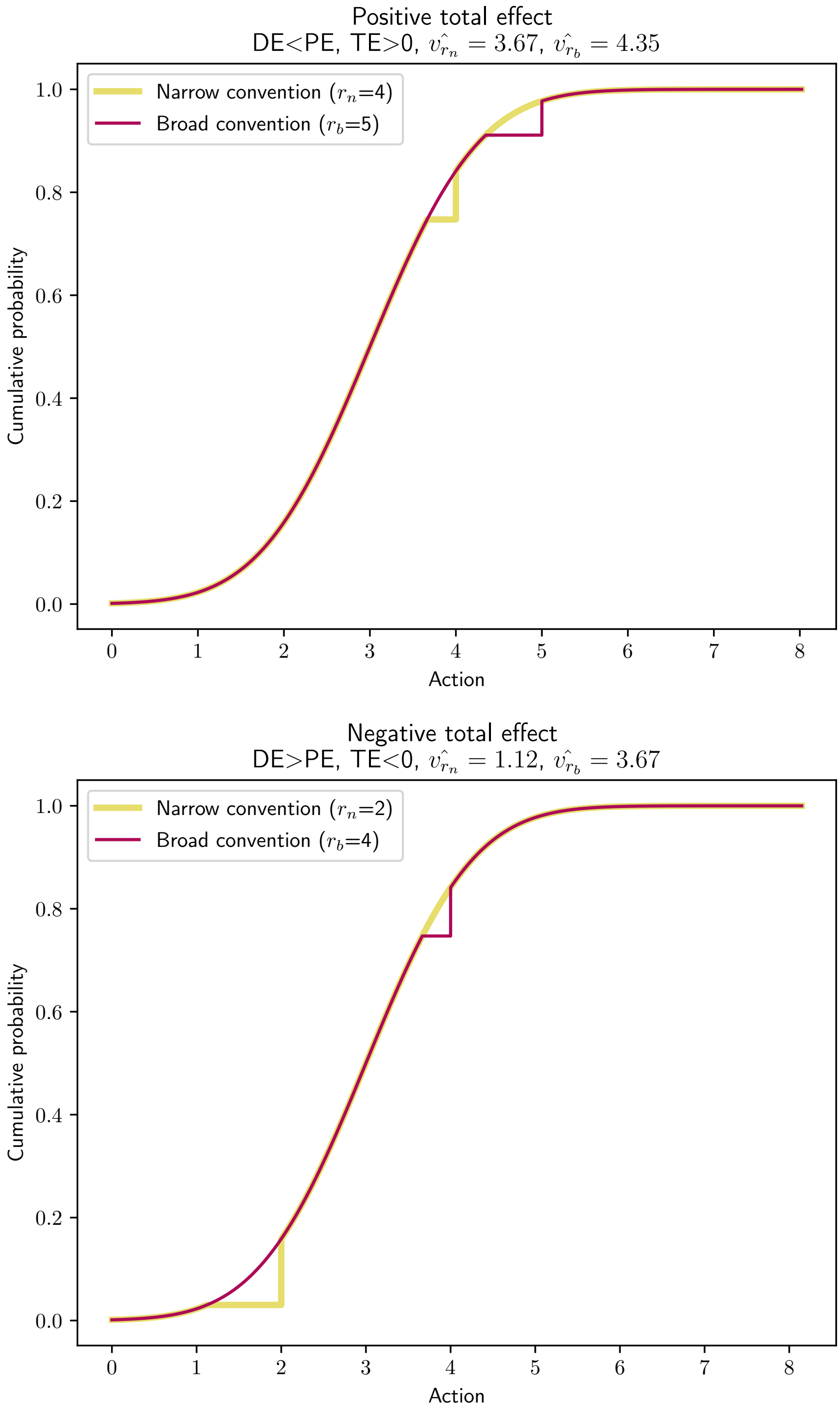


Figure A.1: Ambiguous Total Effect of Moving from Narrow Convention to Broad Convention - Simulation

Notes: Types are drawn from $f(v) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(v-3)^2}$, $\gamma = \frac{2}{5}$. Equilibrium actions are characterized by the action profile in Section 1.2. This Figure illustrates the ambiguous effect of widening a negative moral category discussed in Section A.1.2.

A.2 Tables

Table A.1: Balance Table - Intervention

	(1) Narrow convention first		(2) Broad convention first		T-test Difference
Variable	N	Mean/SE	N	Mean/SE	(1)-(2)
Natural action	132	8.947 (0.583)	123	8.455 (0.579)	0.492
Age	132	24.371 (0.493)	123	24.000 (0.412)	0.371
Female	132	0.523 (0.044)	123	0.488 (0.045)	0.035
Practice task 1 (seconds)	132	129.605 (8.229)	123	133.932 (9.168)	-4.327
Practice task 2 (seconds)	132	394.505 (26.253)	123	358.733 (22.028)	35.772

Notes: The value displayed for t-tests are the differences in the means across the groups. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.2: Actions Across Conventions - Dependent Variable: Action

	(1) Action	(2) First action	(3) Second action	(4) Action
Broad convention	-0.173	1.251**	-1.585***	-0.583
	(0.279)	(0.557)	(0.518)	(0.402)
Broad convention first				0.992**
				(0.466)
Broad convention X broad convention first				0.852
				(0.556)
Natural action	0.792***	0.777***	0.815***	0.796***
	(0.0355)	(0.0418)	(0.0371)	(0.0354)
Constant	2.156***	1.813***	2.473***	1.642***
	(0.365)	(0.480)	(0.439)	(0.431)
Observations	510	255	255	510
R^2	0.586	0.574	0.622	0.598

Notes: Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The dependent variable action can take values 0,1,...,20 and corresponds to the number of real effort tasks a subject chooses to solve. Table A.3 adds control variables.

Table A.3: Actions Across Conventions with Controls - Dependent Variable: Action

	(1) Action	(2) First action	(3) Second action	(4) Action
Broad convention	-0.173	1.297**	-1.610***	-0.583
	(0.280)	(0.554)	(0.514)	(0.403)
Broad convention first				1.027**
				(0.461)
Broad convention X broad convention first				0.852
				(0.557)
Natural action	0.797***	0.783***	0.820***	0.802***
	(0.0362)	(0.0422)	(0.0383)	(0.0360)
Female	0.594	0.694	0.611	0.653
	(0.476)	(0.556)	(0.533)	(0.465)
Age	0.0210	0.0500	0.00261	0.0263
	(0.0524)	(0.0681)	(0.0476)	(0.0543)
Constant	1.305	0.180	2.069*	0.611
	(1.303)	(1.689)	(1.246)	(1.333)
Observations	510	255	255	510
R^2	0.588	0.578	0.624	0.601

Notes: Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The dependent variable action can take values 0,1,...,20 and corresponds to the number of real effort tasks a subject chooses to solve. Table A.2 presents the results without control variables.

Table A.4: Actions Across all Conditions - Dependent Variable: Action

	(1) Action	(2) Action	(3) Action	(4) Action	(5) Action	(6) Action
Narrow convention	0.345 (0.261)	0.345 (0.261)				
Broad convention			0.173 (0.310)	0.173 (0.311)		
Convention					0.259 (0.250)	0.259 (0.251)
Female		-1.010 (0.745)		-0.924 (0.828)		-0.820 (0.776)
Age		0.0203 (0.0587)		-0.00862 (0.0707)		0.00953 (0.0652)
Constant	8.710*** (0.411)	8.729*** (1.583)	8.710*** (0.411)	9.385*** (1.876)	8.710*** (0.411)	8.894*** (1.734)
Observations	510	510	510	510	765	765
R^2	0.001	0.008	0.000	0.004	0.000	0.004

Notes: Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The dependent variable action can take values 0,1,...,20 and corresponds to the number of real effort tasks a subject chooses to solve.

Table A.5: Diluting Effect - Benchmarks

	Expected frequency (p_0)	Experiment frequency	Binomial test *p-value*
Noise around natural action from set			
$\{-1, 0, 1\}$	0.04	0.27	< 0.01
$\{-2, -1, 0, 1, 2\}$	0.05	0.27	< 0.01
$\{-3, -2, -1, 0, 1, 2.3\}$	0.06	0.27	< 0.01
Uniform noise around natural action (restricted to movements)	0.25	0.46	< 0.01

Notes: The first set of tests compares the observed relative frequency of the diluting effect with the relative frequency that would arise if actions under either convention were randomly distributed around the natural action. For each subject, I simulate a narrow convention action, a_n, and a broad convention action, a_b, by independently adding uniformly distributed noise to her natural action. I repeat this simulation $10,000$ times and record the relative frequency of the diluting effect. The average relative frequency of the diluting effect across all repetitions serves as the relative frequency of the diluting effect under the null hypothesis, p_0. Using the two-sided binomial test, I test whether the observed relative frequency of the diluting effect is significantly different from p_0, the simulated relative frequency. I use three different noise sets for this test. The second test restricts the analysis to subjects whose actions differ across the two conventions, i.e., subjects whose movements could be consistent with the diluting effect. For each subject, I simulate a narrow convention action, a_n, and a broad convention action, a_b, with independent draws from the uniform distribution over the interval $[0, 20]$. I repeat this simulation $10,000$ times and record the relative frequency of the diluting effect. The average relative frequency of the diluting effect across all repetitions serves as the relative frequency of the diluting effect under the null hypothesis, p_0. Using the two-sided binomial test, I test whether the observed relative frequency of the diluting effect is significantly different from p_0, the simulated relative frequency.

Table A.6: Purifying Effect - Benchmarks

	Expected frequency (p_0)	Experiment frequency	Binomial test *p-value*
Noise around natural action from set			
$\{-1, 0, 1\}$	0.003	0.25	< 0.01
$\{-2, -1, 0, 1, 2\}$	0.02	0.25	< 0.01
$\{-3, -2, -1, 0, 1, 2.3\}$	0.02	0.25	< 0.01
Uniform noise around natural action (restricted to movements)	0.15	0.43	< 0.01

Notes: The first set of tests compares the observed relative frequency of the purifying effect with the relative frequency that would arise if actions under either convention were randomly distributed around the natural action. For each subject, I simulate a narrow convention action, a_n, and a broad convention action, a_b, by independently adding uniformly distributed noise to her natural action. I repeat this simulation $10,000$ times and record the relative frequency of the purifying effect. The average relative frequency of the purifying effect across all repetitions serves as the relative frequency of the purifying effect under the null hypothesis, p_0. Using the two-sided binomial test, I test whether the observed relative frequency of the purifying effect is significantly different from p_0, the simulated relative frequency. I use three different noise sets for this test. The second test restricts the analysis to subjects whose actions differ across the two conventions, i.e., subjects whose movements could be consistent with the purifying effect. For each subject, I simulate a narrow convention action, a_n, and a broad convention action, a_b, with independent draws from the uniform distribution over the interval $[0, 20]$. I repeat this simulation $10,000$ times and record the relative frequency of the purifying effect. The average relative frequency of the purifying effect across all repetitions serves as the relative frequency of the purifying effect under the null hypothesis, p_0. Using the two-sided binomial test, I test whether the observed relative frequency of the purifying effect is significantly different from p_0, the simulated relative frequency.

Table A.7: Diluting Effect

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Broad convention first	0.0340			0.0403			
	(0.168)			(0.171)			
Natural action		-0.0116					
		(0.0136)					
Natural action < 11			0.556***	0.550***			
			(0.194)	(0.197)			
Image concerns				0.103			
				(0.116)			
Female				0.0457			
				(0.176)			
Age				0.00528			
				(0.0164)			
Belief on signaling value w.r.t							
Share					0.00922***		
					(0.00336)		
Natural action						0.0271	
						(0.0248)	
Social appropriateness							0.0198
							(0.0897)
Constant	-0.628***	-0.512***	-1.008***	-1.329***	-0.799***	-0.662***	-0.608***
	(0.117)	(0.142)	(0.166)	(0.483)	(0.111)	(0.0971)	(0.0849)
Observations	255	255	255	255	255	255	255

Notes: Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. All columns display results from probit regressions with the binary variable diluting effect $\{0, 1\}$ as the dependent variable. The variable natural action < 11 takes on the value 1 if a subject's natural action is smaller than 11 and 0 else. The variable image concerns relies on a 4-point Likert scale measuring self-reported social image concerns with higher values expressing higher social image concerns. Age is measured in years. Beliefs on signaling value w.r.t to share is constructed from the subject's beliefs on what share of subjects is in the bad moral category. The belief on how many take an action smaller than r_n under the narrow convention is subtracted from the belief on how many take an action smaller than r_b under the broad convention. Beliefs on signaling value w.r.t to natural action are constructed from the subject's beliefs on what the average natural action of subjects is in the bad moral category. The belief on the average natural action under the narrow convention is subtracted from the belief on the average natural action under the broad convention. Beliefs on signaling value w.r.t to social appropriateness are constructed from the subject's answers to the Krupka and Weber (2013) measure of how socially appropriate each category is. This variable is constructed using a diff-in-diff approach. First, for each convention, the answer to how prosocial being in the bad moral category is subtracted from the answer to how prosocial being in the bad moral category is. Then this difference under the narrow convention is subtracted from the difference under the broad convention.

Table A.8: Purifying Effect

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Broad convention first	0.178 (0.170)			0.139 (0.180)			
Natural action		0.0825*** (0.0143)					
Natural action in $[11, 17)$			0.828*** (0.213)	0.884*** (0.220)			
Image concerns				0.440*** (0.120)			
Female				0.223 (0.182)			
Age				-0.0132 (0.0184)			
Belief on signaling value w.r.t							
Share					0.00136 (0.00337)		
Natural action						0.0367 (0.0235)	
Social appropriateness							-0.0455 (0.0932)
Constant	-0.748*** (0.121)	-1.477*** (0.177)	-0.828*** (0.0979)	-1.395*** (0.510)	-0.685*** (0.107)	-0.730*** (0.0971)	-0.666*** (0.0863)
Observations	255	255	255	255	255	255	255

Notes: Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. All columns display results from probit regressions with the binary variable purifying effect $\{0, 1\}$ as the dependent variable. The variable natural action in $[11, 17)$ takes on the value 1 if a subject's natural action is weakly larger than 11 and smaller than 17. Else, it takes on the value of 0. The variable image concerns relies on a 4-point Likert scale measuring self-reported social image concerns with higher values expressing higher social image concerns. Age is measured in years. Beliefs on signaling value w.r.t to share are constructed from the subject's beliefs on what share of subjects is in the bad moral category. The belief on how many take an action smaller than r_n under the narrow convention is subtracted from the belief on how many take an action smaller than r_b under the broad convention. Beliefs on signaling value w.r.t to natural action are constructed from the subject's beliefs on what the average natural action of subjects is in the bad moral category. The belief on the average natural action under the narrow convention is subtracted from the belief on the average natural action under the broad convention. Beliefs on signaling value w.r.t to social appropriateness are constructed from the subject's answers to the Krupka and Weber (2013) measure of how socially appropriate each category is. This variable is constructed using a diff-in-diff approach. First, for each convention, the answer to how prosocial being in the bad moral category is subtracted from the answer to how prosocial being in the bad moral category is. Then this difference under the narrow convention is subtracted from the difference under the broad convention.

Table A.9: Dilution - Shares

	Narrow Convention ($j = n$)	Broad Convention ($j = b$)	Difference (B.C. - N.C.)	Test of proportions *p-value*
Share $a_{j \in \{n,b\}} < r_{j \in \{n,b\}}$	0.35	0.65	0.3	< 0.01
Share $a_{j \in \{n,b\}} < r_n$	0.35	0.61	0.26	< 0.01
Share $a_{j \in \{n,b\}} < r_b$	0.9	0.65	-0.25	< 0.01

Notes: The shares are compared using the two-sided test of proportions. The first row compares the share of subjects in the bad moral category across conventions. The second row compares the share of subjects who choose an action less than $r_n = 11$ across conventions. The third row compares the share of subjects who choose an action less than $r_b = 17$ across conventions.

Table A.10: Dilution - Natural Action

	Narrow Convention ($j = n$)	Broad Convention ($j = b$)	Difference (B.C. - N.C.)	T-test *p-value*
$E[a^*\|a_{j\in\{n,b\}} < r_{j\in\{n,b\}}]$	3.03	5.58	2.54	< 0.01
$E[a^*\|a_{j\in\{n,b\}} < r_n]$	3.03	5.26	2.22	< 0.01
$E[a^*\|a_{j\in\{n,b\}} < r_b]$	7.62	5.58	-2.05	< 0.01

Notes: Regressions are clustered at the subject level. The first row compares the natural action of subjects who are in the bad moral category across conventions. The second row compares the natural action of subjects who choose an action less than $r_n = 11$ across conventions. The second row compares the natural action of subjects who choose an action less than $r_b = 17$ across conventions.

Table A.11: Beliefs - Accuracy

	Narrow Convention (average, $j = n$)			Broad Convention (average, $j = b$)		
	True value	Avg. belief	T-test *p-value*	True value	Avg. belief	T-test *p-value*
Share $a_{j\in\{n,b\}} < r_{j\in\{n,b\}}$	0.35	0.43	< 0.01	0.65	0.62	0.14
$E[a^*\|a_{j\in\{n,b\}} < r_n]$	3.03	4.52	< 0.01	5.58	6.32	< 0.01

Notes: The first row compares, for each convention separately, the true share of subjects in the bad moral category with the average belief on the share of subjects in the bad moral category. The second row compares, for each convention separately, the true average natural action of subjects in the bad moral category with the average belief on the average natural action of subjects in the bad moral category. See Figure A.10 for graphical representation.

Table A.12: Beliefs - Dilution

	Narrow Convention (average, $j = n$)	Broad Convention (average, $j = b$)	Difference (B.C. - N.C.)	T-test *p-value*
Belief on share $a_{j\in\{n,b\}} < r_{j\in\{n,b\}}$	0.43	0.62	0.19	< 0.01
Belief on $E[a^*\|a_{j\in\{n,b\}} < r_n]$	4.52	6.32	1.8	< 0.01

Notes: The comparisons employ a two-sided paired t-test. The first row compares the belief on the share of subjects in the bad moral category across conventions. Subjects believe that significantly more subjects are in the bad moral category under the broad convention. The second row compares the belief on the average natural action of subjects in the bad moral category across conventions. Subjects believe that subjects who are in the bad moral category under the broad convention have a significantly higher natural action than subjects who are in the bad moral category under the narrow convention.

Table A.13: Social Appropriateness

	Narrow Convention (average, $j = n$)	Broad Convention (average, $j = b$)	Difference (B.C. - N.C.)	T-test *p-value*
Perceived social appropriateness of bad category	0.97	1.31	0.34	< 0.01
Perceived social appropriateness of good category	2.43	2.63	0.2	< 0.01
Difference (good category - bad category)	1.46	1.33		
T-test *p-value*	< 0.01	< 0.01		

Notes: Using the method by Krupka and Weber (2013), participants rated the social appropriateness of each moral category under each convention on a 4-point Likert scale (0: very antisocial, 1: somewhat antisocial, 2: somewhat prosocial, 4: very prosocial). The first row compares the social appropriateness of the bad moral category across both conventions. The second row compares the social appropriateness of the good moral category across both conventions. The third row compares the appropriateness of each category within a convention, and the fourth row displays the p-value of the respective two-sided paired t-test. For an across-convention graphical representation, see Figure A.12.

Table A.14: Balance Table - Forecasting Experiment

	(1) Forecast narrow convention		(2) Forecast broad convention		T-test Difference
Variable	N	Mean/SE	N	Mean/SE	(1)-(2)
Natural action	99	6.798 (0.739)	99	7.899 (0.753)	-1.101
Female	99	0.717 (0.045)	99	0.727 (0.045)	-0.010
Age	82	24.598 (0.541)	84	23.643 (0.429)	0.955

Notes: The value displayed for t-tests are the differences in the means across the groups. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.15: Comparability of Samples between Experiments

	(1) Main experiment		(2) Forecasting experiment		T-test Difference
Variable	N	Mean/SE	N	Mean/SE	(1)-(2)
Natural action	255	8.710 (0.411)	198	7.348 (0.528)	1.361**
Female	255	0.506 (0.031)	198	0.722 (0.032)	-0.216***
Age	255	24.192 (0.323)	166	24.114 (0.345)	0.078
Climate behavior	255	0.878 (0.021)	198	0.783 (0.029)	0.096***
Climate behavior others	255	0.961 (0.012)	198	0.924 (0.019)	0.037*
Climate behavior belief	255	72.396 (1.035)	198	76.348 (1.264)	-3.952**
Climate norms belief	255	82.243 (0.865)	198	86.455 (0.951)	-4.211***
Environmental care	255	1.757 (0.041)	198	1.606 (0.047)	0.151**

Notes: The value displayed for t-tests are the differences in the means across the groups. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Subjects in the main experiment seem to care more about the environment (i.e. higher natural action, higher self-reported measure of how much they care (environmental care), more agreement to the statement that they try to contribute to climate protection (climate behavior), that others should do that (climate behavior others)). The variables climate behavior, climate behavior others, climate behavior belief and climate norms belief are measured by the survey items proposed in Andre et al. (2021). In contrast, subjects in the forecasting experiment believe more that others are trying to contribute to climate protection (climate behavior belief). They also believe that more subjects agree to the statement that others should contribute to climate protection (climate norms belief). The share of females is higher in the forecasting experiment. Since for the forecasting experiment the age data is pulled from the laboratory's database, some entries are missing. Most of the differences are statistically significant but are generally small in size and possibly affected by preceding elements of the experiments.

Table A.16: Effects - Forecasting

	Frequency			Size			
	Actual	Perceived	T-test *p-value*	Actual	Perceived	T-test *p-value*	Share underestimation
Diluting effect	0.27	0.17	< 0.01	1.51	1.26	< 0.01	0.76
Purifying effect	0.25	0.32	< 0.01	1.35	2.12	< 0.01	0.33

Notes: All t-tests are two-sided. The first row compares the actual frequency (size) of the diluting effect with the perceived frequency (size) of the diluting effect. The size of the diluting effect corresponds to the average decrease in actions due to behavior that is consistent with the diluting effect. The second row compares the actual frequency (size) of the purifying effect with the perceived frequency (size) of the purifying effect. The size of the purifying effect corresponds to the average increase in actions due to behavior that is consistent with the purifying effect.

Table A.17: Effects - Forecasting

	Ratio			
	Actual	Perceived	T-test *p-value*	Share underestimation
DE/PE	1.06	1.2	0.68	0.66
DE/PE (excluding outliers)	1.06	0.63	< 0.01	0.74
PE/DE	0.94	5.91	< 0.01	0.31
PE/DE (excluding outliers)	0.94	3.73	< 0.01	0.41

Notes: All t-tests are two-sided. The table compares the ratio $\frac{DE}{PE}$ and the inverse ratio $\frac{PE}{DE}$ with their perceived counterparts. For each participant, the perceived ratio is constructed based on her perceptions of the frequencies of both effects. Outliers are defined as the lowest 5 percent of ratios and the highest 5 percent of ratios.

Table A.18: Perceptions - Effects

	(1) Perception diluting effect	(2) Perception diluting effect	(3) Perception purifying effect	(4) Perception purifying effect
Forecast broad convention	-0.152***	-0.156***	0.175***	0.179***
	(0.0202)	(0.0203)	(0.0222)	(0.0218)
Natural action	-0.00156		0.000223	
	(0.00140)		(0.00154)	
Diluting effect propensity		0.0325		
		(0.0441)		
Purifying effect propensity				0.0345
				(0.0579)
Sophistication (beauty contest)	-0.0321	-0.0353	0.0587*	0.0594*
	(0.0300)	(0.0299)	(0.0329)	(0.0324)
Female	0.00865	0.0000288	0.0471*	0.0473*
	(0.0231)	(0.0227)	(0.0253)	(0.0250)
Constant	0.277***	0.270***	0.148***	0.143***
	(0.0335)	(0.0340)	(0.0368)	(0.0362)
Observations	198	197	198	197
R^2	0.244	0.244	0.272	0.290

Notes: Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Both the perception of the diluting effect and the perception of the purifying effect are continuous variables between 0 and 1. Based on the data from the main experiment, I construct a propensity score for both effects (diluting effect propensity/purifying effect propensity). It is the conditional likelihood of behaving consistent with the diluting (purifying) effect based on the natural action. The variable sophistication corresponds to the subject's answer in an unincentivized version of the beauty contest. This coarse measure of sophistication is lower for more sophisticated subjects.

A.3 Figures

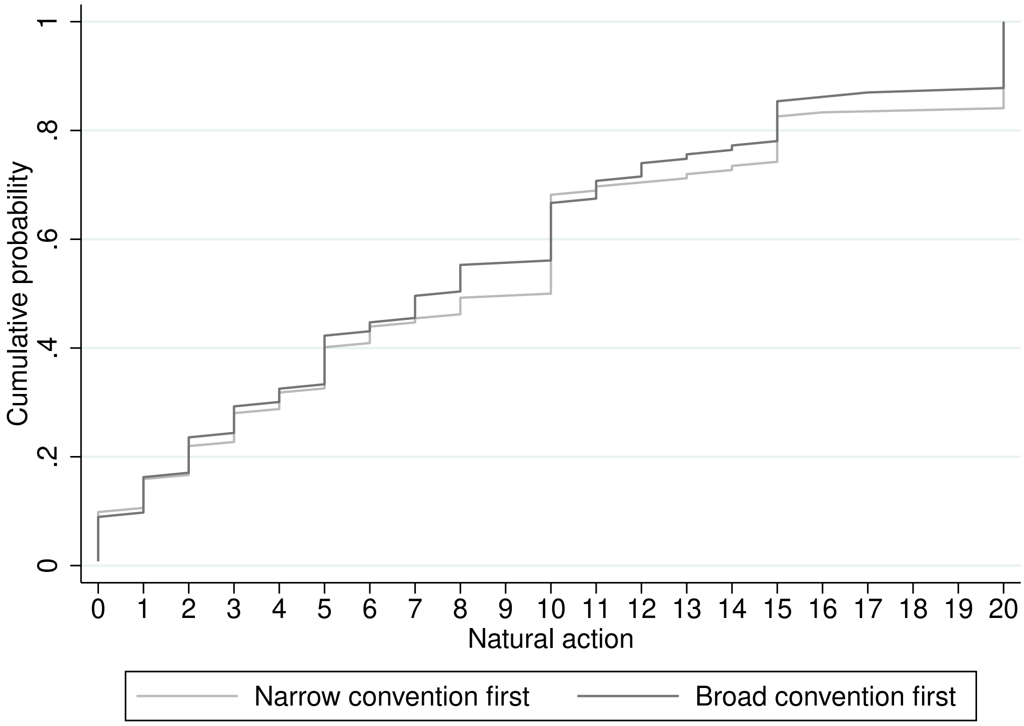


Figure A.1: Natural Action - Balance

Notes: The distribution of natural actions does not differ by the order in which subjects encounter the narrow and the broad convention (p-value of Kolmogorov-Smirnov test = 0.97).

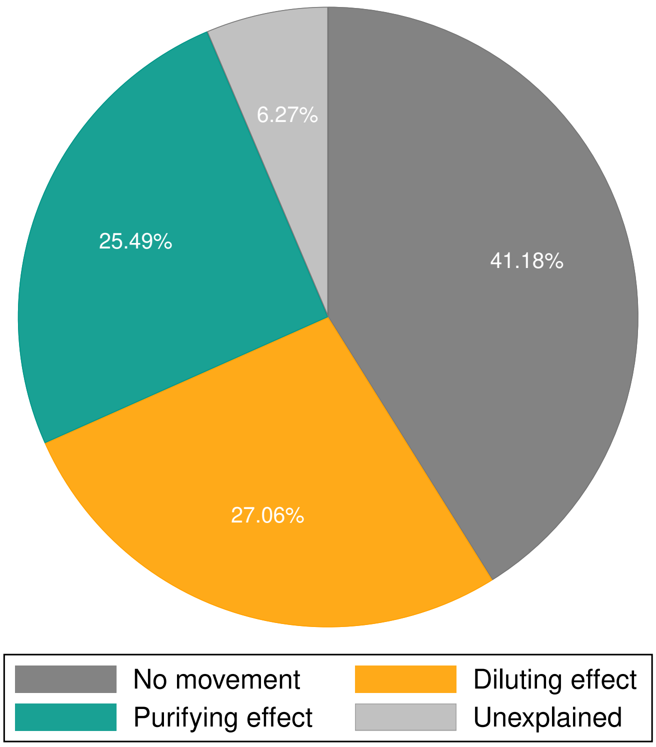


Figure A.2: Movements across Conventions within Subject

Notes: A subject's actions under the Narrow Convention and the Broad Convention may be the same (no movement), consistent with the diluting effect ($a_n \geq r_n = 11 \wedge a_b < r_n = 11$), or consistent with the purifying effect ($a_n < r_b = 17 \wedge a_b \geq r_b = 17$). They may also not fall into any of these categories (unexplained).

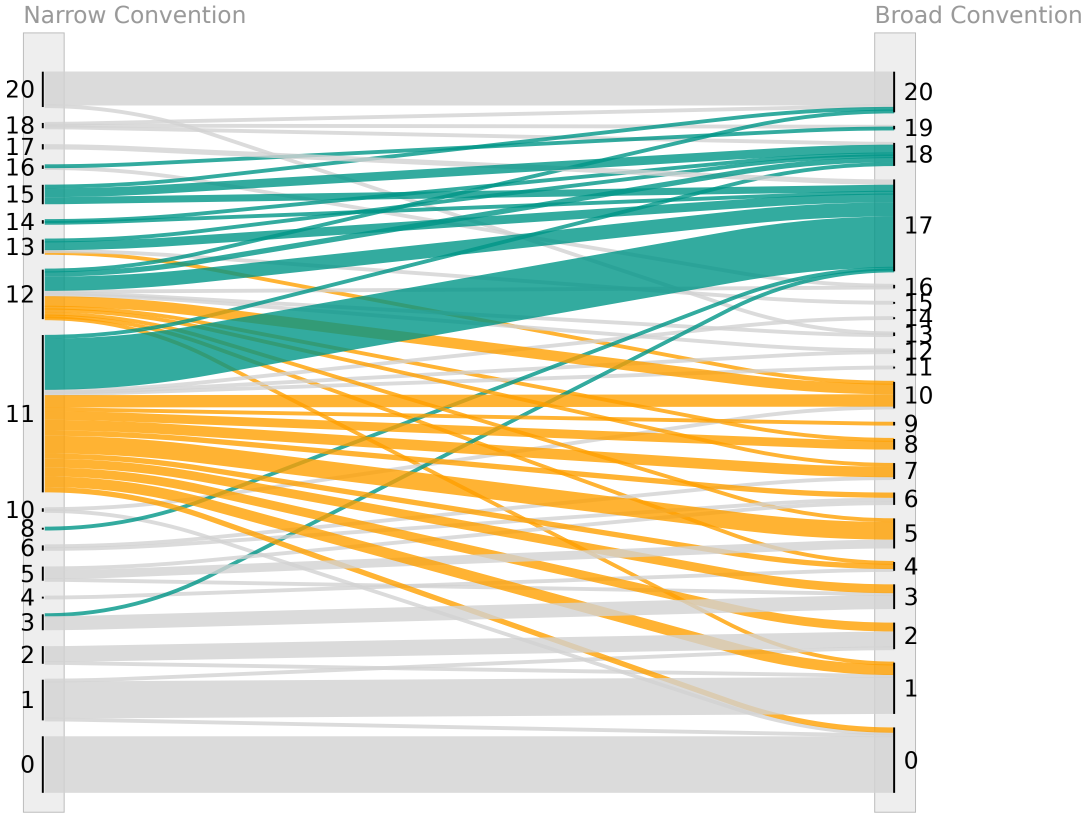


Figure A.3: Movements across Conventions - Diluting Effect and Purifying Effect

Notes: This Sankey graph depicts subjects' actions across the two conventions. A subject's actions under the Narrow Convention and the Broad Convention may be consistent with the diluting effect ($a_n \geq r_n = 11 \wedge a_b < r_n = 11$) or the purifying effect ($a_n < r_b = 17 \wedge a_b \geq r_b = 17$). They may also be inconsistent with any of the two. This graph pools the data across the order in which subjects encountered the two conventions. The breadth of a line reflects the relative frequency of a movement.

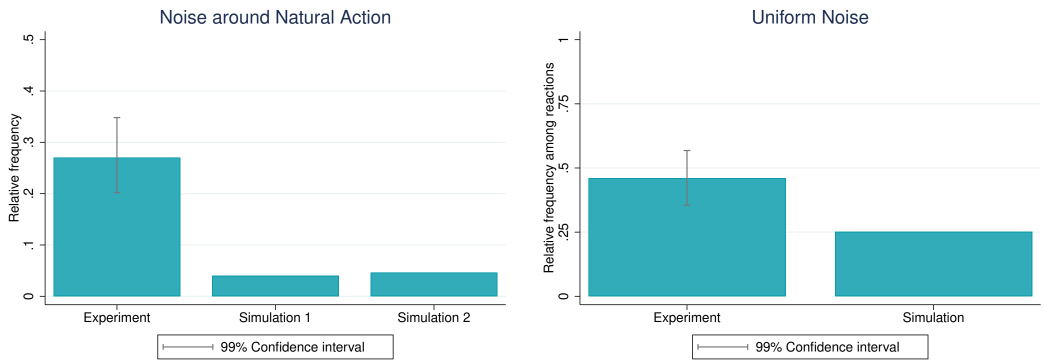


Figure A.4: Diluting Effect

Notes: The left panel compares the observed relative frequency of the diluting effect with the relative frequency that would arise if actions under either convention were randomly distributed around the natural action. For each subject, I simulate a narrow convention action, a_n, and a broad convention action, a_b, by independently adding uniformly distributed noise to her natural action. I repeat this simulation $10,000$ times and record the relative frequency of the diluting effect. The average relative frequency of the diluting effect across all repetitions serves as the relative frequency of the diluting effect under the null hypothesis, p_0. Using the two-sided binomial test, I test whether the observed relative frequency of the diluting effect is significantly different from p_0, the simulated relative frequency. Simulation 1 runs this test by uniformly drawing noise from the set $\{-1, 0, 1\}$, and simulation 2 runs this test by uniformly drawing noise from the set $\{-2, -1, 0, 1, 2\}$. The right panel restricts the analysis to subjects whose actions differ across the two conventions, i.e., subjects whose movements could be consistent with the diluting effect. For each subject, I simulate a narrow convention action, a_n, and a broad convention action, a_b, with independent draws from the uniform distribution over the interval $[0, 20]$. I repeat this simulation $10,000$ times and record the relative frequency of the diluting effect. The average relative frequency of the diluting effect across all repetitions serves as the relative frequency of the diluting effect under the null hypothesis, p_0. Using the two-sided binomial test, I test whether the observed relative frequency of the diluting effect is significantly different from p_0, the simulated relative frequency. The exact frequencies are reported in Table A.5.

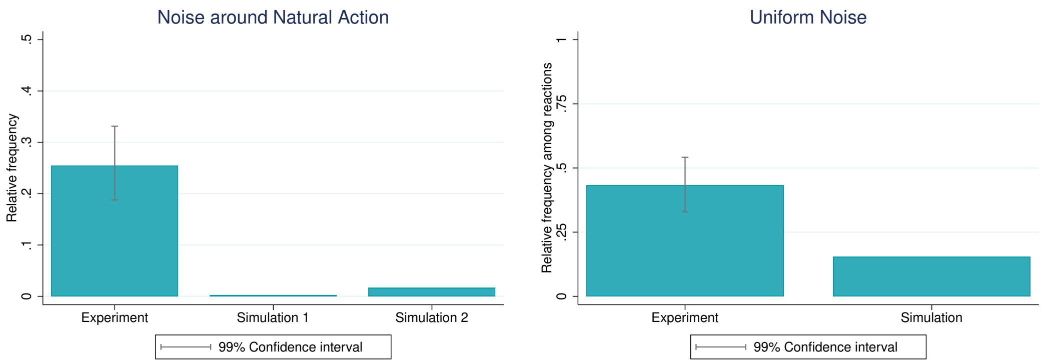


Figure A.5: Purifying Effect

Notes: The left panel compares the observed relative frequency of the purifying effect with the relative frequency that would arise if actions under either convention were randomly distributed around the natural action. For each subject, I simulate a narrow convention action, a_n, and a broad convention action, a_b, by independently adding uniformly distributed noise to her natural action. I repeat this simulation $10,000$ times and record the relative frequency of the purifying effect. The average relative frequency of the purifying effect across all repetitions serves as the relative frequency of the purifying effect under the null hypothesis, p_0. Using the two-sided binomial test, I test whether the observed relative frequency of the purifying effect is significantly different from p_0, the simulated relative frequency. Simulation 1 runs this test by uniformly drawing noise from the set $\{-1, 0, 1\}$, and simulation 2 runs this test by uniformly drawing noise from the set $\{-2, -1, 0, 1, 2\}$. The right panel restricts the analysis to subjects whose actions differ across the two conventions, i.e., subjects whose movements could be consistent with the purifying effect. For each subject, I simulate a narrow convention action, a_n, and a broad convention action, a_b, with independent draws from the uniform distribution over the interval $[0, 20]$. I repeat this simulation $10,000$ times and record the relative frequency of the purifying effect. The average relative frequency of the purifying effect across all repetitions serves as the relative frequency of the purifying effect under the null hypothesis, p_0. Using the two-sided binomial test, I test whether the observed relative frequency of the purifying effect is significantly different from p_0, the simulated relative frequency. The exact frequencies are reported in Table A.6.

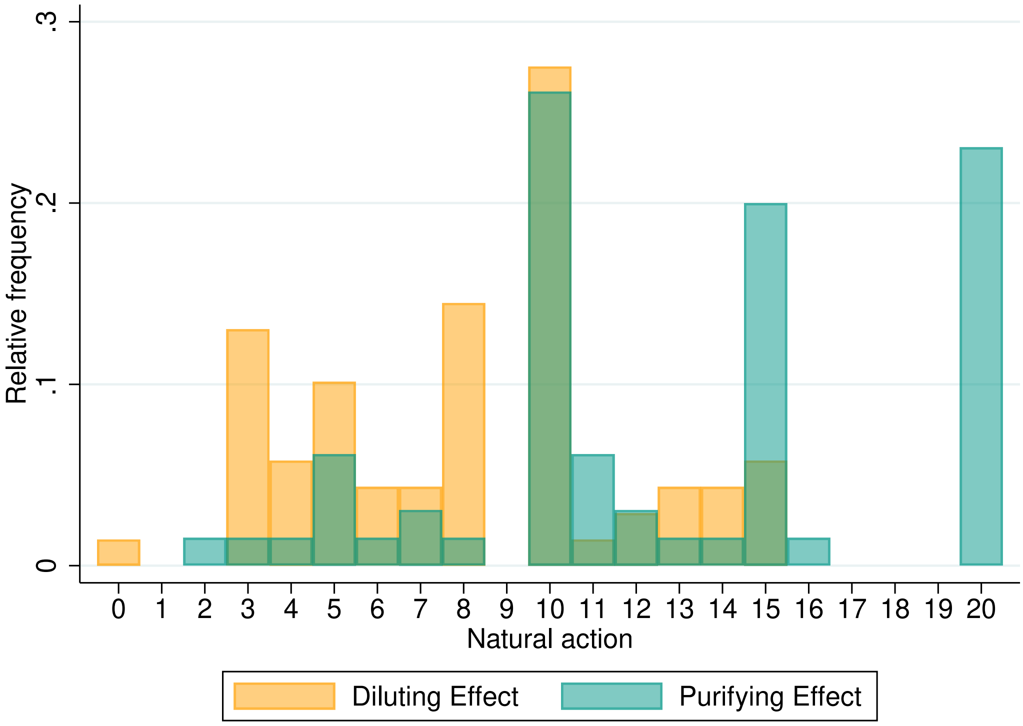


Figure A.6: Natural Action Conditional on Diluting and Purifying Effect

Notes: This figure depicts (i) the histogram of natural actions of subjects whose actions are consistent with the diluting effect ($a_n \geq r_n = 11 \wedge a_b < r_n = 11$) and (ii) the histogram of natural actions of subjects whose actions are consistent with the purifying effect ($a_n < r_b = 17 \wedge a_b \geq r_b = 17$).

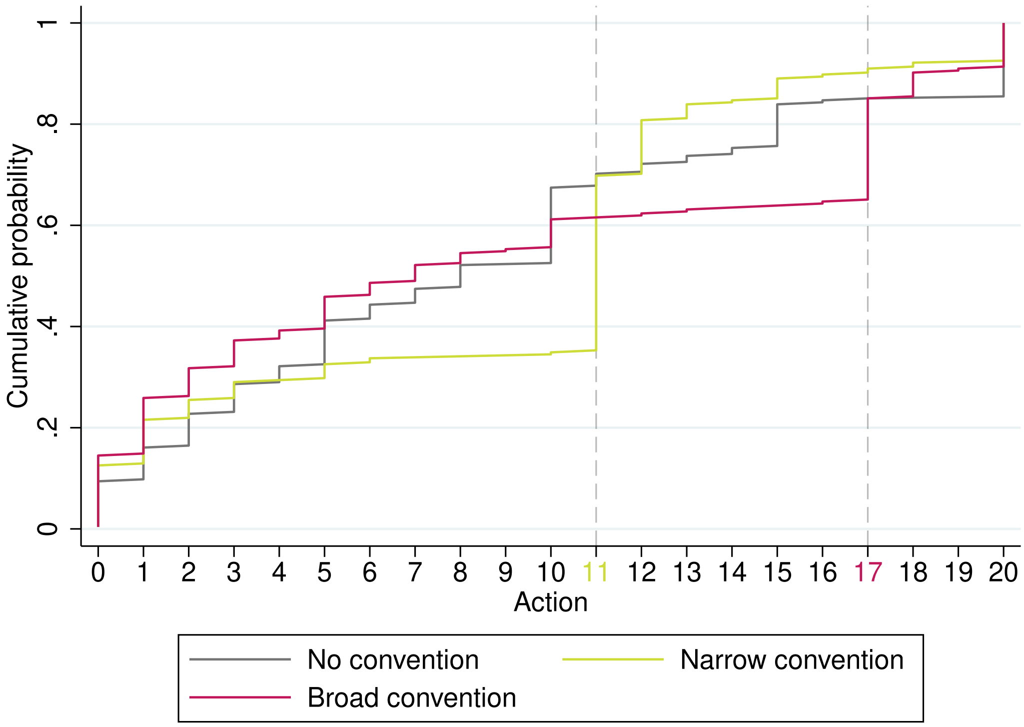


Figure A.7: Cumulative Distribution of Actions under All Conventions

Notes: The threshold actions of both conventions are marked (i.e. $r_n = 11$, $r_b = 17$). No convention corresponds to the Natural Action Condition.

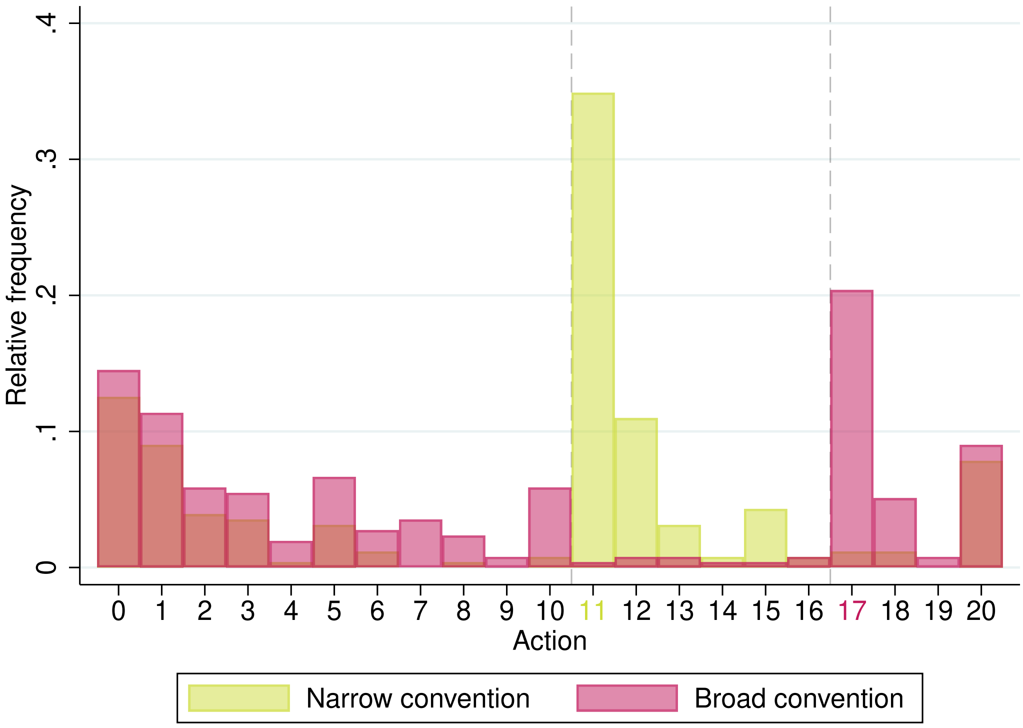


Figure A.8: Histogram of Actions under Conventions

Notes: The threshold actions of both conventions are marked (i.e. $r_n = 11$, $r_b = 17$).

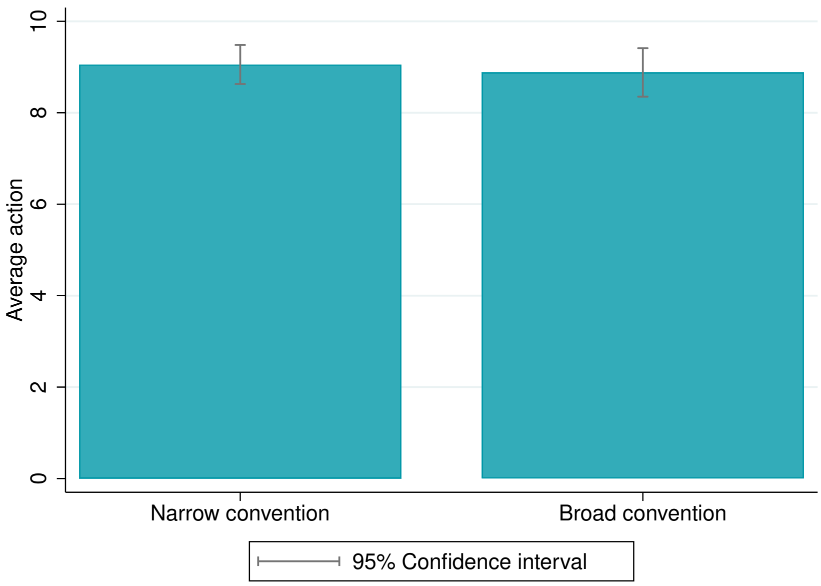


Figure A.9: Average Action under Narrow and Broad Convention

Notes: The average action under the narrow convention (9.05, $r_n = 11$) is not statistically different from the average action under the broad convention (8.88, $r_b = 17$).

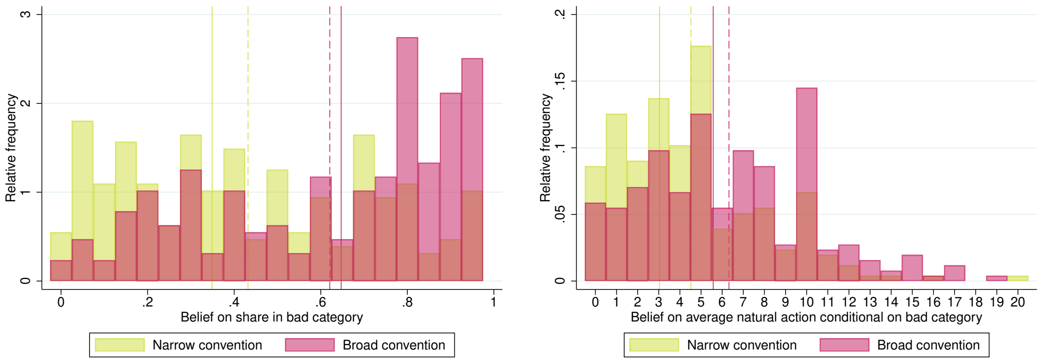


Figure A.10: Beliefs -Accuracy

Notes: The solid line represents the true average and the dashed line represents the average estimate. The left panel displays the distribution of beliefs on the share of subjects in the bad moral category under both conventions. The right panel displays the distribution of beliefs on the average natural action conditional on the bad category under both conventions. See Table A.11 for exact values and statistical tests.

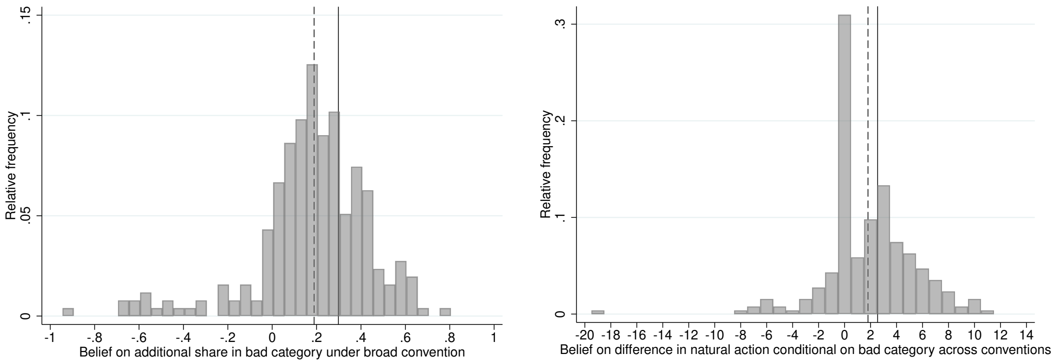


Figure A.11: Perceived Shift in Meaning

Notes: The solid line represents the true value and the dashed line represents the average estimate. The left panel shows that there is a 30 percentage points higher share of subjects in the bad moral category under the broad convention than under the narrow convention (solid line). On average, subjects believe there only to be an additional 18 percentage points of subjects in the bad moral category under broad convention (dashed line) (p-value of two-sided t-test < 0.01). Using this metric, 65% of subjects underestimate the extent to which the bad category is diluted. Similarly, the right panel shows that the natural action conditional on the bad category is higher by 2.54 under broad convention than under narrow convention (solid line). On average, subjects believe the natural action conditional on the bad category to be higher only by 1.8 under the broad convention (dashed line) (p-value of two-sided t-test < 0.01; robust to excluding the outlier). Using this metric, 60% of subjects underestimate the extent to which the bad category is diluted.

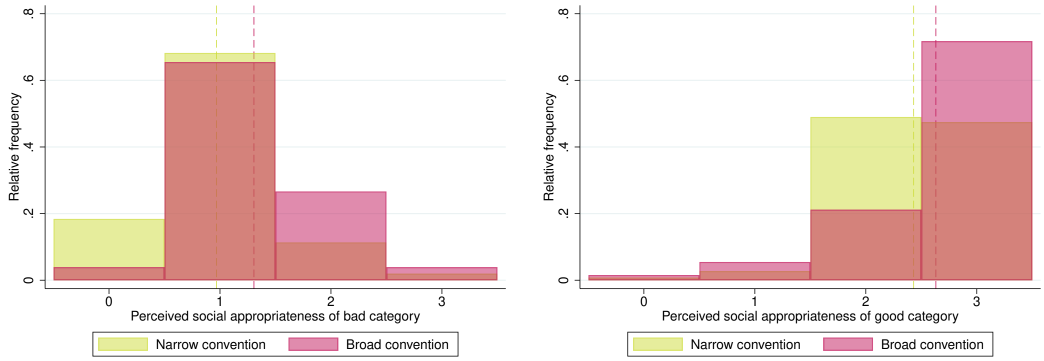


Figure A.12: Social Appropriateness

Notes: The dashed lines represent the average answers. Using the method by Krupka and Weber (2013), participants rated the social appropriateness of each moral category under each convention on a 4-point Likert scale (0: very antisocial, 1: somewhat antisocial, 2: somewhat prosocial, 4: very prosocial). See Table A.13 for exact values and t-test.

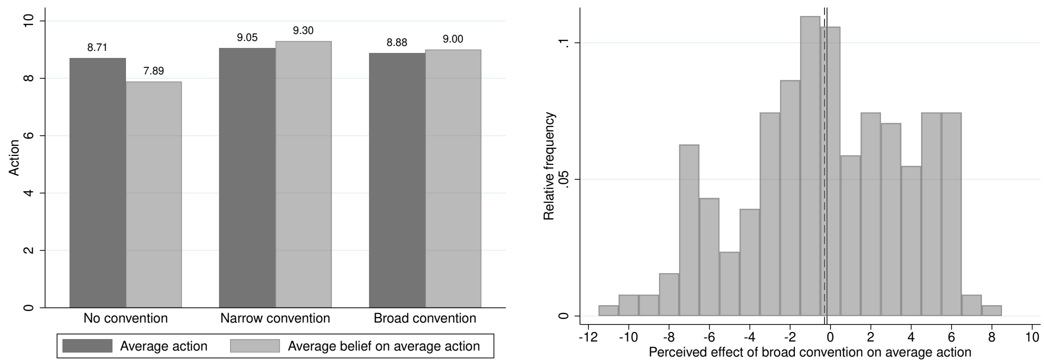


Figure A.13: Beliefs on Effect of Conditions on Actions

Notes: The left panel depicts the average action across all three conditions and the average belief on the average action across all three conditions. Table A.4 shows that there are no statistically significant differences across conditions. Subjects incorrectly believe, on average, that either convention will cause significantly higher actions than no convention (p-value of two-sided t-test < 0.01 for both comparisons). Subjects correctly believe that the average contribution does not differ across the narrow and the broad convention (p-value of two-sided t-test $= 0.25$). The right panel depicts the implied beliefs on the average treatment effect of the broad convention. Subjects only stated their beliefs on the average action under each convention. To get the perceived effect of the broad convention on the average action I subtract the belief on the average action under the narrow convention from the belief on the average action under the broad convention. The solid line represents the true effect (-0.17) and the dashed line represents the average estimated effect (-0.29). Perceptions of the effect are not statistically different from the true effect (p-value of two-sided t-test $= 0.63$). 49% of subjects underestimate the effect of the broad convention on actions.

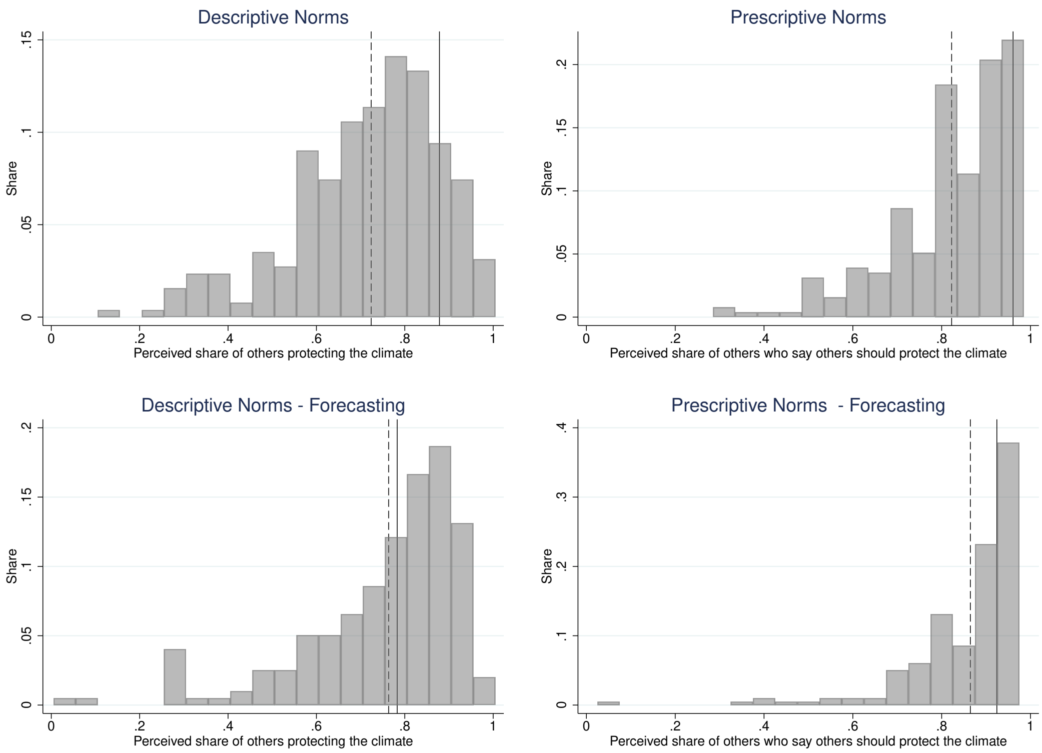


Figure A.14: Social Norms around Climate Action

Notes: The solid line represents the true average and the dashed line represents the average estimate. The upper (lower) panels report data from the main (forecasting) experiment. Using the method by Andre et al. (2021), subjects indicated whether they are contributing to climate protection (Yes/No). They were then incentivized to guess the share of other participants who said yes to the previous question (descriptive norms). In a second question, subjects indicated whether they think others should contribute to climate protection (Yes/No). They were then incentivized to guess the share of other participants who said yes to the previous question (prescriptive norms).

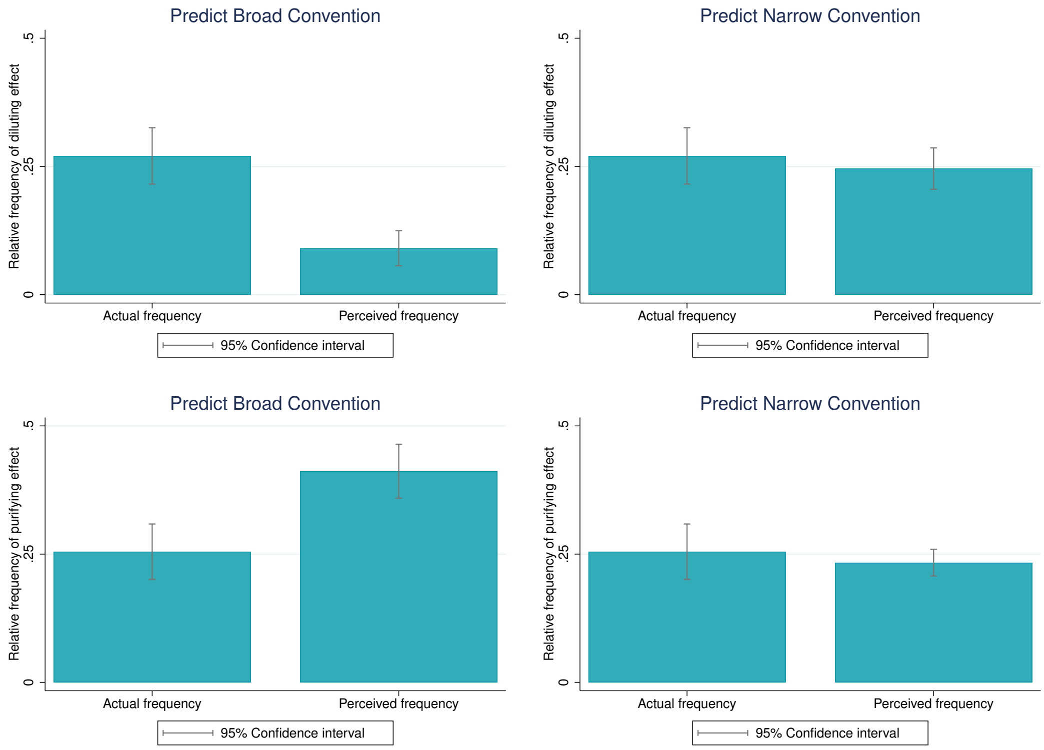


Figure A.15: Perceptions of Diluting and Purifying Effect - Treatment Variation

Notes: The upper (lower) panels display the actual and perceived frequencies of the diluting (purifying) effect split up by treatment. Subjects assigned to the Predict Broad Convention condition perceive 9% (41%) of subjects to behave consistently with the diluting (purifying) effect. Subjects assigned to the Predict Narrow Convention condition perceive 25% (23%) of subjects to behave consistently with the diluting (purifying) effect. In fact, 27% (25%) behaved consistently with the diluting (purifying) effect.

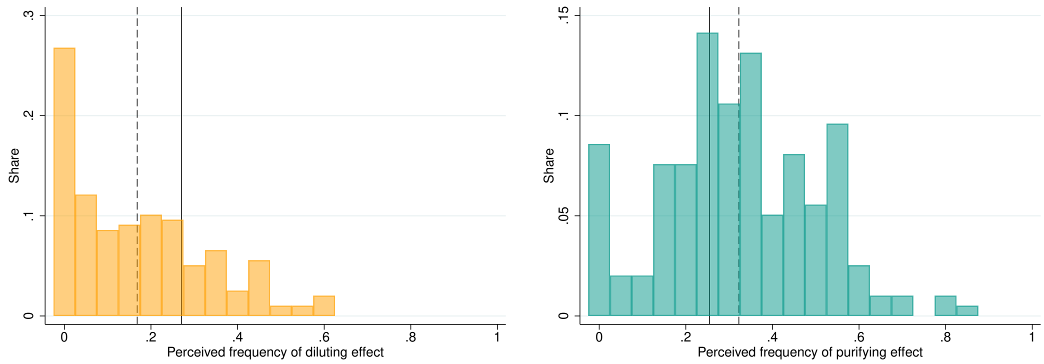


Figure A.16: Heterogeneity in Perceptions of Diluting and Purifying Effect

Notes: In the left panel, the solid line represents the true frequency of the diluting effect (0.27), and the dashed line represents the average perceived frequency of the diluting effect (0.16). In the right panel, the solid line represents the true frequency of the purifying effect (0.25), and the dashed line represents the average perceived frequency of the diluting effect (0.32). The figure shows that individual perceptions vary.

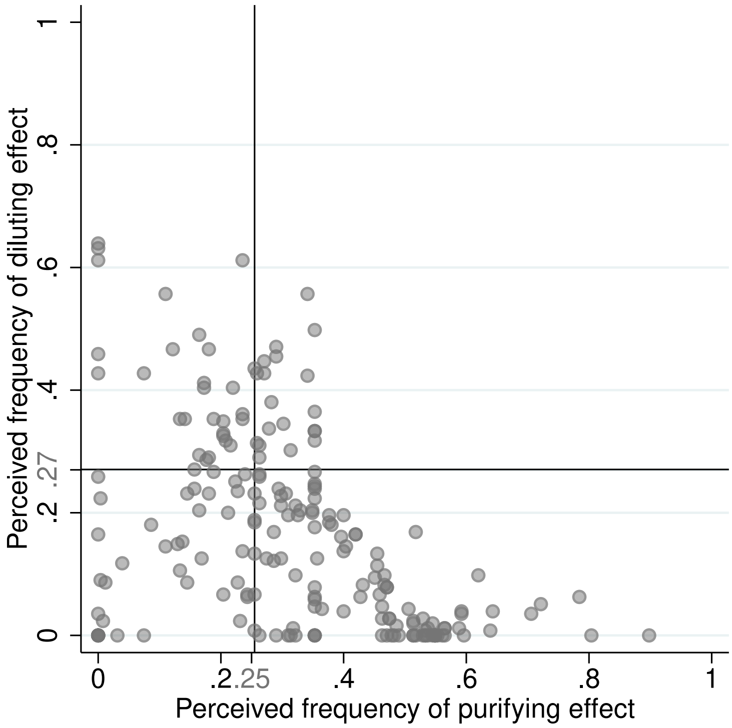


Figure A.17: Joint Distribution of Perceptions of Effects

Notes: The cross represents the actual frequencies of the purifying effect (0.25) and the diluting effect (0.27). A share of 0.1 of subjects' perceptions is northeast of the intersection. A share of 0.57 of subjects' perceptions is southeast of the intersection. A share of 0.19 of subjects' perceptions is southwest of the intersection. A share of 0.14 of subjects' perceptions is northwest of the intersection.

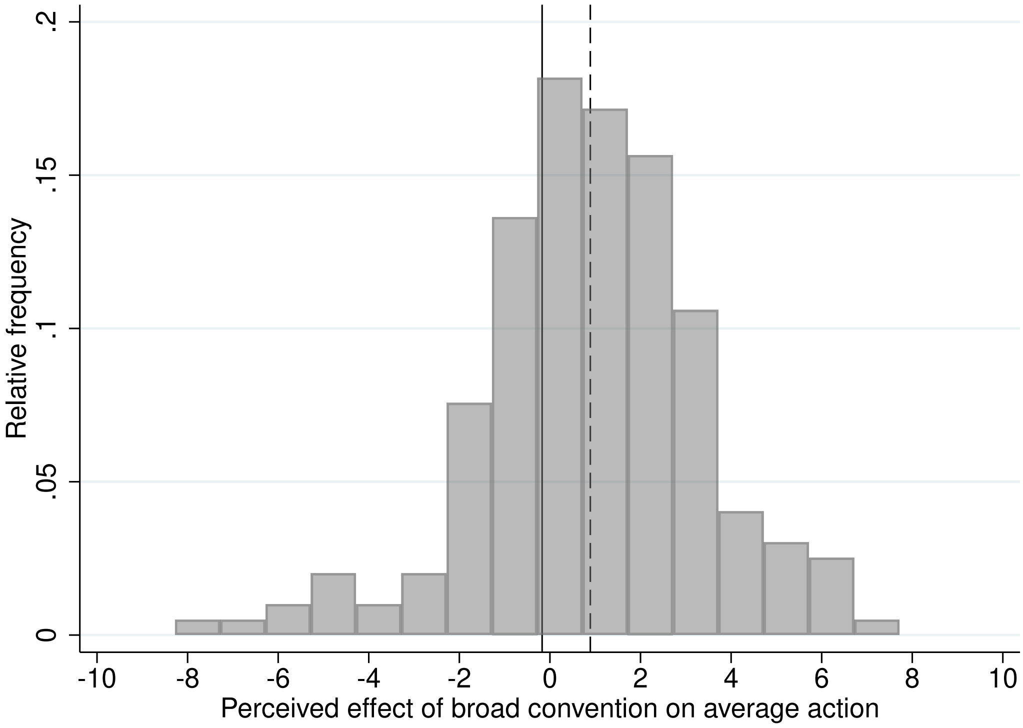


Figure A.18: Perceived Effect of Broad Convention on Average Action

Notes: The solid line represents the true treatment effect of the broad convention, and the dashed line represents the average perceived treatment effect of the broad convention. The perceived treatment effect of the broad convention is constructed from the subjects' stated perceptions of how participants in the main experiment behaved under each convention. The average perceived average treatment effect is 0.89, which is larger than the actual average treatment effect of -0.17 (p-value of two-sided t-test < 0.01). This overestimation of the effect of the broad convention on the average action is entirely driven by subjects assigned to the Predict Broad Convention condition.

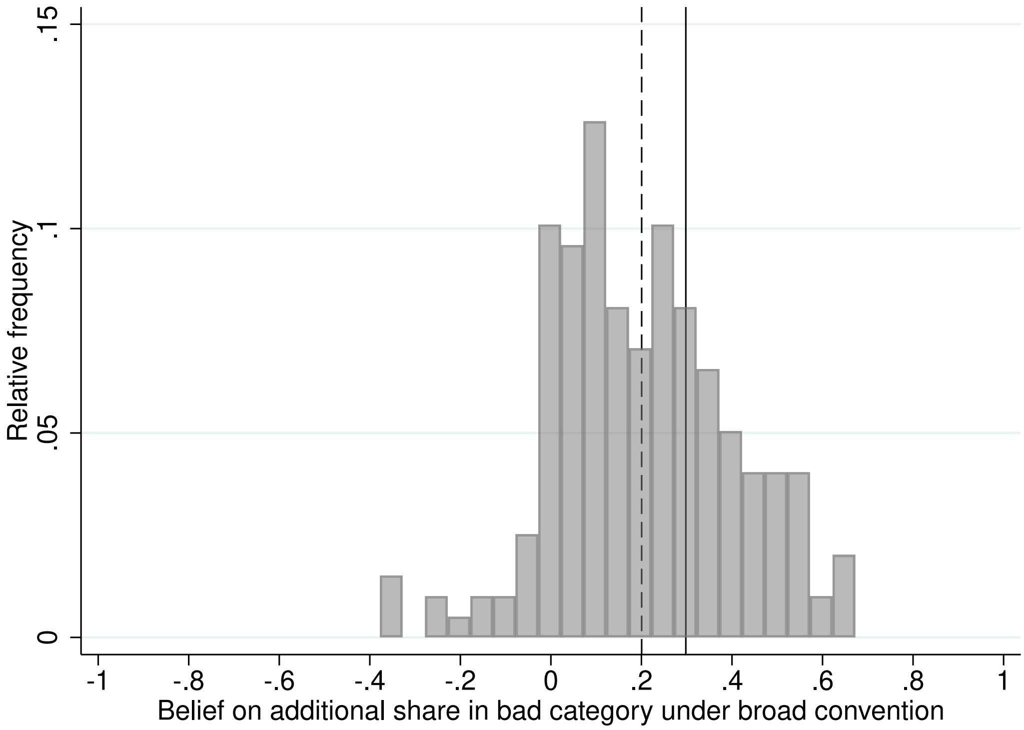


Figure A.19: Perceived Shift in Meaning - Forecasting

Notes: The solid line represents the true value, and the dashed line represents the average estimate. The figure shows that there is a 30 percentage points higher share of subjects in the bad moral category under the broad convention than under the narrow convention (solid line). On average, subjects believe there only to be an additional 20 percentage points of subjects in the bad moral category under broad convention (dashed line) (p-value of two-sided t-test < 0.01). 55% of subjects underestimate the extent to which the bad category is diluted. These numbers are similar to the numbers obtained from the original experiment (see Figure A.11, left panel).

A.4 Experimental Material

A.4.1 Instructions - Main Experiment

I Instructions - Baseline

Consent

When conducting this survey, data will be generated by the decisions you have to make. This data will be scientifically analyzed by researchers at LMU Munich. In the process, the decision data is anonymized and cannot be traced back to any individual. In this sense, participation in the survey is anonymous. The anonymized data generated will be used for the preparation of scientific research papers and presentations and, if necessary, made available to other researchers. These papers will be published. As part of this experiment, your choices can partly be observed by other students in Munich. In this case, the other students will see you via Zoom while your camera is switched on and you are logged in with your participant ID, and they will receive information about your decisions. By giving your consent, you confirm that you agree to the use of your decision-making data as described above and that you wish to participate in the survey.
Reply options:

- *I consent, begin the study.*
- *I do not consent, I do not wish to participate.*

——— *new page* ———

Please enter your participant ID:

——— *new page* ———

Pronoun

Which pronoun should be used to refer to you?
Reply options:

- *He*
- *She*

——— *new page* ———

Welcome

Welcome to this experiment, and thank you for your participation.
In this experiment, you will make choices on your computer. Depending on your choices, you can earn extra money. Your payment for today's experiment consists of the following parts:

- Show-up fee of **6 euros**

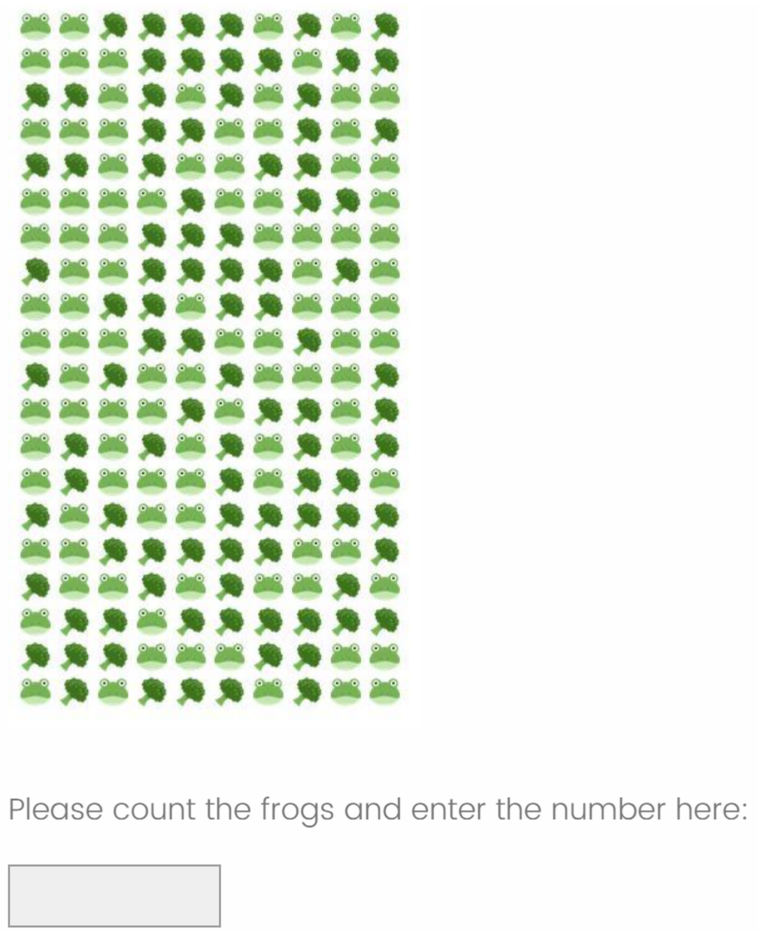


Figure A.20: Example of Real Effort Task with 20 Lines

Participants who chose 11 tasks under condition A

- 89 participants choose 11 tasks under condition A.
- They were all "protecting the climate" under condition A.
- Under condition B they would be "protecting the climate" if they chose 17 tasks or more.
- In condition B, the participants could decide on exactly the same number, more or fewer tasks than under condition A.

How many tasks did the 89 participants choose to protect under condition B?

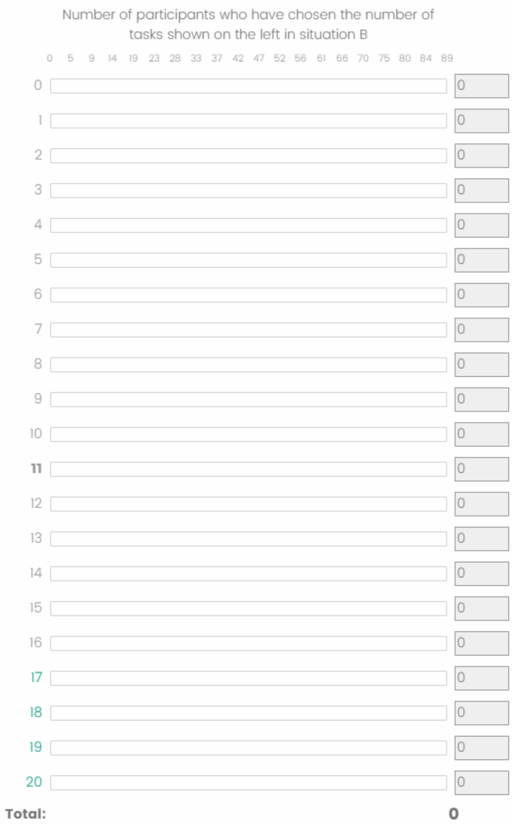


Figure A.21: Elicitation of Beliefs - Forecasting Experiment

Notes: This is an example of 1 of the 21 decision screens that subjects who were assigned to the Predict Broad Convention condition faced in the Forecasting Experiment.

- **Additional 12 euros** for completing the experiment
- Additional money depending on your choices

Please now put aside your mobile phone and close any other browser windows or email programs. If you have any **questions** during the experiment, please **write them to "MELESSA" via Zoom chat**. Your question will be answered in the chat.
In the following, you will find detailed information about your participation in this experiment. This information is the same for all participants. Afterward, you will answer some **comprehension questions**.

——— *new page* ———

The consequences of your decisions
In today's experiment, you and the other participants will make decisions. These decisions have consequences, which will be described in more detail later.

Consequences
In today's experiment, you will make a decision in **three different situations**. One of these situations will be chosen at random and your decision in that situation will be implemented. Each of these three situations will be chosen with the same probability. **So, since each of your decisions could actually be implemented, you should make your decision carefully**.

——— *new page* ———

Information on using moorlands for climate protection
Here is information that is relevant to your decision. It is about the use of moorlands for climate protection. Please read the following text and graphics carefully.
Moors in Germany have the potential to avoid the release of **several million tonnes of carbon dioxide (CO2)** each year. If drained moors are rewetted, they can counteract climate warming as a natural climate solution.
How much carbon do moorlands store?
<info-graphic: moors occupy only a small fraction of the land surface but store twice as much carbon as the total biomass of all forests combined (500 billion tons of CO2)>
How do drained moorlands differ from rewetted moorlands?
<graphic illustration of drained moorlands emitting carbon dioxide and rewetted moorlands storing carbon>
How can we help to use moorlands for climate protection?
The registered association BUND Naturschutz in Bayern has set itself the task of re-wetting

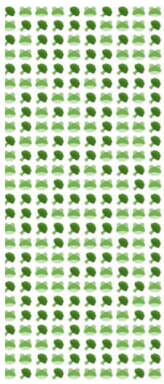
drained moorlands in Bavaria. Every euro donated enables the purchase of one square meter of moorland for rewetting and, thus, for climate protection.

new page

Decision

When making your decision, you have **in all three situations** the option to provide up to one square meter for the rewetting of moorland. You can choose how many of a maximum of 20 tasks you would like to solve. For each task solved, 5 eurocents will be donated to BUND Naturschutz e.V. (Friends of the Earth Germany). Every euro donated enables the purchase of one square meter of moorland for rewetting and thus for **climate protection (1 task solved:5 dm^2 (square decimeter) of moorland).**

What does a task look like?



Please type the number of frogs that you can count below: (This is only an example, you do not have to actually solve the task.)

new page

What is my choice?

You decide how many of the maximum of 20 tasks you would like to solve: none at all, a few, or all 20 tasks.

- With each task you solve, you **make a contribution to protecting the climate** (1 task solved:5 dm^2 of moorland). CO_2 can be stored in the protected moorland area.
- With every task you do not solve, **you do not contribute to climate protection**. CO_2 can escape from the unprotected moorland area. The fewer tasks you solve, the **sooner** you **finish the experiment**.

Your payment for today's experiment does not depend on how many tasks you

You do not solve any task.	You do not protect any moorland for the climate.
You solve 12 tasks.	You protect 60 dm^2 of moorland for the climate.
You solve all 20 tasks.	You protect 100 dm^2 ($1m^2$) of moorland for the climate.

solve.

new page

Which of my choices will be implemented?

You choose how many tasks you want to solve in all three situations. How the situations differ from each other is described later. After you have chosen for each situation how much you want to contribute to climate protection (i.e. how many tasks you want to solve), one situation is chosen at random. **You then have to solve the number of tasks you have committed to in the randomly selected situation**.

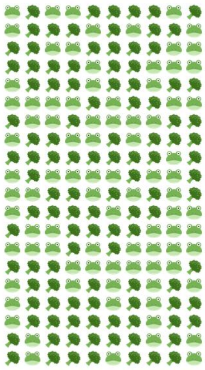
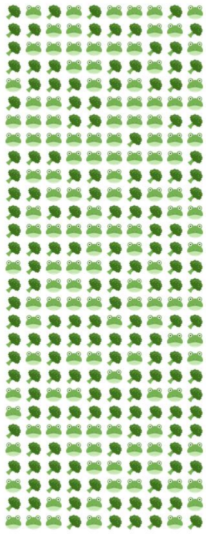
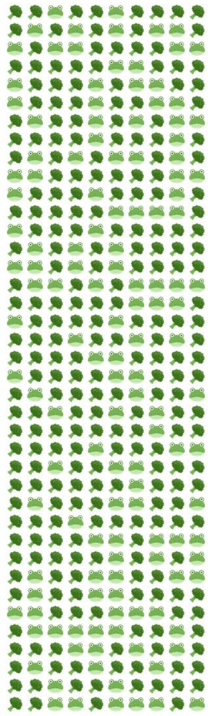
new page

Tasks

The twenty tasks are of different lengths, so they vary in difficulty. The tasks will progressively become more difficult. The first task is 20 lines long, the second task is 21 lines long, ..., the nineteenth task is 38 lines long, and the twentieth task is 39 lines long. For a task with 20 lines, participants need about 2 minutes. For a task with 39 lines, students need about 6 minutes. For all 20 tasks, students need, on average, one hour.

Examples for tasks - please do not solve

To solve a task, you must count and enter the number of frogs on the screen.

A first task (20 lines)	A tenth task (29 lines)	A twentieth task (39 lines)
		

(Guessing is not worthwhile because there is a time interval during which input is not possible if a wrong answer is entered.)

new page

The three situations

Everything that has been described so far applies to all three situations. You also have all the knowledge you need to make your decision (how many tasks to solve) for situation 1. You will find out later how situations 2 and 3 differ from situation 1.

The rest of the procedure is as follows:

- Summary of instructions
- Comprehension questions
- Practice round – you familiarize yourself with the tasks
- You make your choice for situation 1
- Description of situation 2
- Description of situation 3
- Comprehension questions
- You make your choice for situation 2

- You make your choice for situation 3
- One of the situations is randomly chosen
- Questionnaire
- You solve the number of tasks set for this situation
- Payment: You receive the link for the payment

We will ask you at several points about your assessments (for example, about the behavior of the other participants). These questions are an important part of the experiment. The more correct you are with your assessments, the more extra money you can earn.

new page

Comprehension questions

We now ask you to answer some questions about the content of the experiment. You can continue with the experiment only if you correctly answer all the questions.

Summary

Here is a summary of the most important information.

Using moorland for climate protection

- Drained, unprotected moorlands can leak CO2. CO2 can be stored in rewetted, protected moors.

Your decision

- You choose how many of a maximum of 20 tasks you want to solve.
- You choose this decision for each of the three situations.
- After you have decided how many tasks you would like to solve in each situation, one situation is chosen at random.
- You must solve the number of tasks you set for that situation.
- For each task you solve, 5 eurocents will be donated to BUND Naturschutz e.V. (Friends of the Earth Germany). Every euro donated enables the purchase of one square meter of moorland for rewetting and thus for **climate protection (1 task solved:5 dm^2 of moorland).**
- A task looks like this: <grid of frogs and broccoli emojis>
- Your payment for today's experiment does not depend on how many tasks you solve.

Tasks

- The tasks become increasingly difficult.
- The later tasks are longer (more lines) than the first tasks.

Situations

- **Everything described above applies to all three situations.**
- You now have all the knowledge you need to make your decision (how many tasks to solve) for situation 1.
- Situations 2 and 3 will be explained later.

This summary was available to participants while answering the comprehension questions. This option will be represented by the button:

☐ Display summary

new page

Comprehension question 1

Your payment for today's experiment does not depend on how many tasks you solve.

Reply options:

- *True.*
- *False.*

☐ Display summary

new page

Comprehension question 2

The ____ tasks you solve, the more moorland will be protected.

Reply options:

- *less*
- *more*

☐ Display summary

new page

Comprehension question 3

The tasks will become progressively ____.

Reply options:

- *shorter.*
- *longer.*
- *The tasks stay the same length.*

☐ Display summary

——— *new page* ———

Comprehension question 4

Assume that for situation 1, you have chosen to solve 5 tasks. For situation 2 and situation 3, you have also chosen how many tasks you want to solve. Then, situation 1 is chosen at random. How many tasks do you have to solve?

☐ Display summary

——— *new page* ———

Practice tasks

Thank you. You have now answered all the questions correctly. You will now work through 2 practice exercises to familiarize yourself with the task. Afterward, you will choose how many of the maximum of 20 tasks you would like to solve in each situation.

——— *new page* ———

Practice task 1/2 (20 lines)

<grid of frogs and broccoli emojis with 20 lines>

Please count the frogs and enter their number here:

(The timer above shows you how long this task takes you. After 5 seconds, you can move on to the next screen).

——— *new page* ———

Practice task 2/2 (39 lines)

<grid of frogs and broccoli emojis with 39 lines>

Please count the frogs and enter their number here:

(The timer above shows you how long this task takes you. After 5 seconds, you can move on to the next screen).

——— *new page* ———

Thank you, you have now solved all the practice tasks. On the next screen, you will choose how many of the maximum 20 tasks you want to solve in Situation 1.

If you still have questions, now is a good time to ask them to “MELESSA” via the chat.

☐ I have no further questions and want to continue.

——— *new page* ———

II Natural Action Condition

Your decision for situation 1

How many tasks do you want to solve?

<slider from 0 to 20>

- 1 task solved:5 dm^2 of protected moorland.
- One of the three situations (situation 1, situation 2, situation 3) is chosen at random.
- Each situation is selected with the same probability.
- You have to solve the number of tasks you have set for the randomly selected situation.
- The tasks will progressively become more difficult.
- The fewer tasks you solve, the sooner you are finished with the experiment and can leave

new page

III Instructions – Social Judgement

Situation 2 and situation 3

Thank you for your choice in situation 1. Situations 2 and 3 are similar to each other and are based on situation 1. The following still applies:

- 1 solved task:5 dm^2 of moorland.
- One of the three situations (situation 1, situation 2, situation 3) is chosen at random.
- Each situation is chosen with the same probability.
- You have to solve the number of tasks you have set for the randomly selected situation.
- The tasks will become progressively longer.

In situation 2 and situation 3, the other participants can partly see your decision.

new page

Situation 2 and situation 3

In situation 2 and situation 3, the other participants can partly see your decision.

What do the other participants see about your decision?

The other participants see whether you **solve more or less than a certain number of tasks -** i.e. whether you protect more or less than a certain area of moorland. **This number of tasks is different in situation 2 and situation 3.** Let's now call this number x (x is different in situation 2 and situation 3).

How will the other participants know whether you have solved more or less than x tasks?

After you and the other participants have decided how many tasks you solve in situation 2 and situation 3, one situation is randomly chosen out of the three situations. If situation 2 or situation 3 is randomly selected, all participants will be called up in turn by their participant ID.

You solve less than x tasks

Suppose situation 2 is randomly selected, and you have previously decided to solve less than x tasks in situation 2. When you are called up, the following will be read out:

"<Participant ID> is not protecting the climate because he/she is solving less than x tasks, protecting less than x*5 (calculated beforehand) dm^2 (square decimeters) of moorland for the climate."

Also, you will add "not protecting the climate" to your Zoom name beforehand, so that your Zoom name is:

<Participant ID> not protecting the climate

(Your renaming will be checked by MELESSA. If you rename yourself incorrectly, it will be corrected).

During your turn, your video will be spotlighted, which means it will be clearly visible to all other participants.

You solve x tasks or more tasks

Suppose situation 2 is randomly selected, and you have previously decided to solve x tasks or more tasks in situation 2. When you are called up, the following will be read out:

"<Participant ID> is protecting the climate because he/she is solving x tasks or more, protecting x*5 (calculated beforehand) dm^2 (square decimeters) or more of moorland for the climate."

Also, you will add "protecting the climate" to your Zoom name beforehand, so your Zoom name is:

<Participant ID> protecting the climate

During your turn, your video will be spotlighted, which means it will be clearly visible to all other participants.

Here is an example of what the Zoom meeting looks like when it is the turn of the participant with participant ID 582 and it is read out that he is not protecting the climate:
<graphic illustration of spotlighting and renaming>

new page

Situation 2 and situation 3

What else do the other participants know other than that you have solved more or less than x tasks?

None of the other participants will know exactly how many tasks you solve. No one will know if you solve a different number of tasks in different situations. After a situation has been randomly chosen, but before the decisions of all participants are made partially visible, it will be announced how many participants are protecting the climate and how many participants are not protecting the climate.

new page

Assessments in situation 2 and situation 3

Therefore, the other participants will know if you are "protecting the climate" or "not protecting the climate". All participants will think about what it means in situation 2 and what it means in situation 3 not to be "protecting the climate" or to be "not protecting the climate". They will do this by answering the questions below.

Remember: To be "protecting the climate" in situation 2, you have to solve a different number of tasks than in situation 3 (x is different). Otherwise, both situations are exactly the same.

The questions here are only for illustration - please do not answer them. You can only answer them when you know x for situation 2 and situation 3.

How many tasks do you think someone solved in **situation 1** (when nothing was observed) who was ...

- ... in Situation 2 (x=?) "not protecting the climate": <slider 0,1...,20>
- ... in Situation 3 (x=?) "not protecting the climate": <slider 0,1...,20>

If you are less than 1 task away from the true value, you get **an extra 0.50 euros**.

How much does someone care about the climate who is ...

- ... "not protecting the climate" in Situation 2 (x=?): *very little, little, rather much, very much*
- ... "not protecting the climate" in Situation 3 (x=?): *very little, little, rather much, very much*

——————— *new page* ———————

Comprehension questions

We now ask you to answer some questions about the content of the experiment. You can only continue with the experiment if you correctly answer all the questions. Here is a summary of the main information once again.

Summary

What still applies:

Your decision

- You choose how many of a maximum of 20 tasks you want to solve.
- You make this decision in all three situations.
- After you have decided how many tasks you want to solve in each situation, one situation will be chosen at random. You must solve the number of tasks you committed to for that situation.
- For each task that you solve, 5 eurocents will be donated to BUND Naturschutz e.V. (Friends of the Earth Germany). Every euro donated enables the purchase of one square meter of moorland for rewetting and thus for climate protection (1 task solved:5 dm^2 of moorland).
- A task looks like this: <grid of frogs and broccoli emojis>
- Your payment for today's experiment does not depend on how many tasks you solve.

Tasks

- The tasks get harder over time.
- The later tasks are longer (more lines) than the first tasks.

What's new:

- In situation 2 and situation 3, the other participants can partly see your decision.
- The other participants can see whether you solve **more or less than a certain number of tasks (x).**

What do the other participants see?

- If you do less than x tasks, the other participants see that you are "not protecting the climate".
- If you solve x tasks or more than x tasks, the other participants see that you are "protecting the climate".

How do situation 2 and situation 3 differ from each other?

- This number of tasks (x) that you have to solve to be "protecting the climate" is different in situation 2 and situation 3.
- So, depending on the situation, there is a difference in how difficult it is to be "protecting the climate".

new page

Comprehension question 1

What do the other participants see in situation 2 and situation 3?

Reply options:

- *Exactly how many tasks you solve.*
- *Whether you solve more or less than x tasks.*

☐ Display summary

new page

Comprehension question 2

What does it mean that the minimum number of tasks (x) you have to solve to be "protecting the climate" is different in situation 2 and situation 3?

Reply options:

- *That the tasks are different in difficulty depending on the situation.*
- *That protecting the climate is different in difficulty depending on the situation.*

☐ Display summary

new page

Thank you, you have now correctly answered all the comprehension questions. In the following, you will decide how many of the maximum 20 tasks you want to solve in Situation 2 and

Situation 3.

If you still have questions, now is a good time to ask them to “MELESSA” via the chat.

☐ I have no further questions and want to continue.

new page

At this stage the threshold for situation 2 was revealed $r_n = 11$ or $r_b = 17$. The instructions below are written for the case where the subject was randomly assigned to the Narrow Convention First condition.

Your assessments for situation 2

In situation 2, you need to solve at least **11 tasks** to be considered to be “protecting the climate”.

Before you choose how many tasks to solve in situation 2 on the next screen, we ask you for some assessments.

Now we do not want to hear your opinion, but **what other participants would say about the decisions below in situation 2.** Your answers will be compared with the answers of the other participants (all students in Munich). In situation 2 there are two possible cases: Someone is protecting the climate or someone is not protecting the climate, depending on whether they solve more or less than 11 tasks. One row of the table below will be randomly selected. If your answer matches what was **chosen most often by the other participants, you will receive an additional 0.50 euros. Please indicate how you think other participants evaluate the possible case.**

- To not be protecting the climate (to solve less than 11 tasks) is ... Reply options: *Very antisocial, Somewhat antisocial, Somewhat prosocial, Very prosocial*
- To be protecting the climate (to solve less than 11 tasks) is ... Reply options: *Very antisocial, Somewhat antisocial, Somewhat prosocial, Very prosocial*

In situation 2, what percentage of participants are not protecting the climate?

(If you have to solve 11 tasks or more to be protecting the climate.)

If you are less than 5 percentage points away from the true value, you will receive an additional 0.50 euros.

<slider 0,1...,100(%)>

new page

Your decision in situation 2

In situation 2, you need to solve at least **11 tasks** to be considered to be “protecting the climate”.

How many tasks do you want to solve?

<slider from 0 to 20>

- 1 task solved:5 dm^2 protected moorland.
- One of the three situations (situation 1, situation 2, situation 3) is chosen at random.
- Each situation is selected with the same probability.
- You have to solve the number of tasks you have committed to in the randomly selected situation.
- The fewer tasks you solve, the sooner you finish the experiment and can leave.
- If you solve fewer than 11 tasks, you are "not protecting the climate".
- If you solve 11 tasks or more, you are "protecting the climate".

new page

At this stage the threshold for situation 3 is revealed $r_n = 11$ or $r_b = 17$. The instructions below are written for the case where the subject was randomly assigned to the Narrow Convention First condition.

Your assessments in situation 3

In situation 3, you need to solve at least **17 tasks** to be considered to be "protecting the climate".

Before you choose how many tasks to solve in situation 3 on the next screen, we ask you for some assessments.

Now we do not want to hear your opinion, but **what other participants would say about the decisions below in situation 3.** Your answers will be compared with the answers of the other participants (all students in Munich). In situation 3, there are two possible cases: Someone is protecting the climate or someone is not protecting the climate, depending on whether they solve more or less than 17 tasks. One row of the table below will be randomly selected. If your answer matches what was **chosen most often by the other participants, you will receive an additional 0.50 euros. Please indicate how you think other participants evaluate the possible case.**

- To not be protecting the climate (to solve less than 17 tasks) is ... Reply options: *Very antisocial, Somewhat antisocial, Somewhat prosocial, Very prosocial*
- To be protecting the climate (to solve less than 17 tasks) is ... Reply options: *Very antisocial, Somewhat antisocial, Somewhat prosocial, Very prosocial*

In situation 3, what percentage of participants are not protecting the climate?

(If you have to solve 17 tasks or more to be protecting the climate.)

If you are less than 5 percentage points away from the true value, you will receive an additional 0.50 euros.

<slider 0,1...,100(%)>

new page

Your decision in situation 3

In situation 3, you need to solve at least **17 tasks** to be considered to be "protecting the climate".

How many tasks do you want to solve?

<slider from 0 to 20>

- 1 task solved:5 dm^2 protected moorland.
- One of the three situations (situation 1, situation 2, situation 3) is chosen at random.
- Each situation is selected with the same probability.
- You have to solve the number of tasks you have committed to in the randomly selected situation.
- The fewer tasks you solve, the sooner you finish the experiment and can leave.
- If you solve fewer than 17 tasks, you are "not protecting the climate".
- If you solve 17 tasks or more, you are "protecting the climate".

new page

Thank you, you have made your choices for all situations. Before the other participants can potentially see whether you are protecting the climate or not and you start with the tasks, we ask you for your **assessments** of the other participants' behavior.

new page

IV Beliefs

Situation 1

How many tasks do participants solve on average?

(If nothing is known about their decision).

If you are less than 1 task away from the true value, you will receive **an additional 0.50 euros.**

<slider from 0 to 20>

Situation 2

How many tasks do participants solve on average?

(If you have to solve 11 amount of tasks or more to be protecting the climate).

If you are less than 1 task away from the true value, you will receive **an additional 0.50 euros.** <slider from 0 to 20>

Situation 3

How many tasks do participants solve on average?

(If you have to solve 17 amount of tasks or more to be protecting the climate).

If you are less than 1 task away from the true value, you will receive **an additional 0.50 euros.** <slider from 0 to 20>

new page

Your assessments

What does "not protecting the climate" mean in situation 2? What does "not protecting the climate" mean in situation 3?

How many tasks do you think someone solved in **situation 1** (when nothing was observed) who was ...

If you are less than 1 task away from the true value, you get **an extra 0.50 euros.**

- ... not protecting the climate in situation 2 (x= 11) <slider from 0 to 20>
- ... not protecting the climate in situation 3 (x= 17) <slider from 0 to 20>

How much does someone care about the environment who is...

- ... not protecting the climate in situation 2 (x= 11) Reply options: *Very little, Rather little, Rather much, Very much*
- ... not protecting the climate in situation 3 (x= 17) Reply options: *Very little, Rather little, Rather much, Very much*

new page

Please wait

Please send your Participant ID via chat to "MELESSA" to notify them that you have arrived here.

Please wait for the other participants. Once everyone is here, the password will be posted in the chat. It will then be communicated which situation has been randomly selected.

Password:

new page

If either situation 2 or 3 is randomly selected, the social judgement part follows. Otherwise, it is skipped. The following is written for the case where the narrow convention (x= 11) is randomly drawn and happens to be situation 2 for the subject. Imagine the subject decided to solve 15 tasks under the narrow convention.

V Social Judgement

Judgement

Situation 2 was selected at random. For this situation, you have chosen 15 tasks to solve. Therefore, you are "protecting the climate" and have to solve 15 tasks afterwards.

Please change your Zoom name to: ID is protecting the climate

(To do this, hover over your name in the Participant Area and click on Rename. Your renaming will be checked by MELESSA. If you rename yourself incorrectly, this will be corrected).

All participants will be called one after the other by their participant ID.

Then, a new password will be posted in the chat. As soon as you have the password and have entered it, you can continue with the rest of the experiment.

Password:

Participants are called up one by and their video is spotlighted while it is announced whether they are protecting the climate or not protecting the climate and what that means for moorland protection.

new page

VI Questionnaire on Norms and Demographics

Remainder of experiment

You can now continue with the experiment. To do so, **please leave the Zoom meeting.** Please contact the email address at the bottom of each screen if you have any questions. At the end of the experiment, you will find the link for payment. You will need to provide your payment details via this link so that we can make the payment. For the payment, you must complete the experiment in full.

new page

Your assessments

Do you try to protect the climate?

Reply options:

- *Yes*
- *No*

Do you think students in Munich should try to protect the climate?

Reply options:

- *Yes*
- *No*

——————————— *new page* ———————————

Your assessments

We are asking the same two questions to many students in Munich. What do you think?

How many of the students in Munich say they try to protect the climate?

If you are less than 5 percentage points away from the true value, you will receive **an additional 0.50 euros.**

<slider 0,1...,100(%)>

How many of the students in Munich say that students in Munich should try to protect the climate?

If you are less than 5 percentage points away from the true value, you will receive **an additional 0.50 euros.**

<slider 0,1...,100(%)>

——————————— *new page* ———————————

How much do you care about the climate?

Reply options:

- *Very little.*
- *Rather little.*
- *Rather much.*
- *Very much.*

——————————— *new page* ———————————

Your assessments

Assume that you have the **goal that the other participants solve as many tasks as possible,** i.e. protect as much moorland as possible.

With this goal in mind, at what point should you be considered to be protecting the climate? How many tasks (out of a maximum of 20 tasks) should you have to solve to

be protecting the climate, so that as many tasks as possible are solved in total?
<slider from 0 to 20>

new page

Thank you very much. Here is a short questionnaire. After that, we will ask you for some personal information. Then you will complete the tasks. After that, you will have finished the experiment.

new page

Questionnaire

How important is it to you what others think of you?
Reply options:

- *Not important at all.*
- *Not very important.*
- *Rather important.*
- *Very important.*

Do you know any of the other participants?
Reply options:

- *Yes*
- *No*

How many of the other participants do you know? Display this question if the answer to the previous question is yes.
Reply options:

- *One.*
- *Two.*
- *More than two.*

How do you personally assess yourself: **Are you generally a risk-taker or do you try to avoid risks?**
Please tick a box on the scale, the value 0 meaning: "not at all willing to take risks" and the value 10 meaning: "very willing to take risks". You can use the values in between to grade your assessment.
Reply options: 0, 1, ..., 10

new page

Do you agree with the following statement?

It is a good small contribution to the environment to solve a task and thus generate a donation for moorland conservation.

Reply options:

- *I fully disagree.*
- *I rather disagree.*
- *I rather agree.*
- *I fully agree.*

What did you think of the frog counting task?

Reply options:

- *Very annoying.*
- *Rather annoying.*
- *Rather pleasant.*
- *Very pleasant.*

Do you think that the fact that others could see their behavior played a role in the other participants' decisions?

Reply options:

- *Yes*
- *No*

How easy to understand do you find this experiment?

Reply options:

- *Very difficult to understand.*
- *Rather difficult to understand.*
- *Rather easy to understand.*
- *Very easy to understand.*

——— *new page* ———

Personal details

How old are you?

What is your gender?

Reply options:

- *female*

- *male*
- *other*
- *I do not wish to answer*

Is German your native language?

Reply options:

- *Yes*
- *No*

What is your major?

new page

VII Real Effort Tasks

Between 0 and 20 such screens, depending on the subject's choice for the randomly selected situation.

<grid of frogs and broccoli emojis>

Please count the frogs and enter their number here:

new page

What did you think of the frog counting task?

Reply options:

- *Very annoying.*
- *Rather annoying.*
- *Rather pleasant.*
- *Very pleasant.*

new page

If there is anything else you would like to tell us, please do so here:

A.4.2 Instructions - Forecasting Experiment

Subjects were randomly assigned to either the Predict Broad Convention condition or to the Predict Narrow Convention condition. For subjects who were assigned to the Predict Broad (Narrow) Convention condition A in what follows corresponds to the Narrow (Broad) Convention Condition of the original experiment, and condition B in what follows corresponds to the Broad (Narrow) Convention Condition of the original experiment. Subjects assigned to the Predict Broad (Narrow) Convention had to predict behavior (i.e. the chosen number of real effort tasks) under the broad (narrow) convention based on the behavior under the narrow (broad) convention. Below are the instructions for the Predict Broad Convention condition. Condition A and condition B are flipped in the Predict Narrow Convention condition. Important screens will also be presented for the Predict Narrow Convention version.

I Instructions

Consent

When conducting this survey, data will be generated by the decisions you have to make. This data will be scientifically analyzed by researchers at LMU Munich. In the process, the decision data is made anonymous and cannot be traced back to any individual. In this sense, participation in the survey is anonymous. The anonymized data generated will be used for the preparation of scientific research papers and presentations and, if necessary, made available to other researchers. These papers will be published. By giving your consent, you confirm that you agree to the use of your decision-making data as described above and that you wish to participate in the survey.

Reply options:

- *I consent, begin the study.*
- *I do not consent, I do not wish to participate.*

——— *new page* ———

Please enter your participant ID:

——— *new page* ———

Welcome

Welcome to this survey. Please be assured that your responses will be kept completely confidential.

The survey should take you around 40 minutes to complete, and you will receive 6 euros for your participation. On top of that and depending on your answers you can earn a 5 euros bonus.

Your participation in this survey is voluntary. You have the right to withdraw at any point.

——— *new page* ———

What is this survey about?

Today, you will be asked to predict the results of a study that we ran a few months ago. In this previous study, we have observed environmentally friendly behavior under different circumstances. Now we are interested in hearing what you think: **How did the circumstances affect environmentally friendly behavior?**

On the following screens, we will briefly summarize the previous study for you, such that you can make good predictions. **The more accurate your predictions, the higher your chances of winning the 5 euros bonus.**

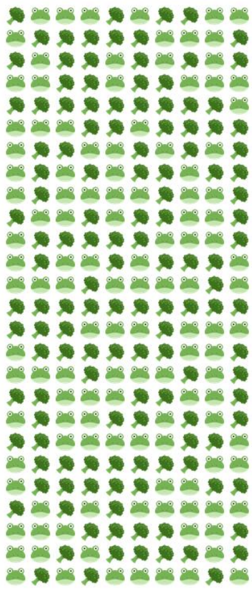
Please note that you cannot go back to a previous page after having proceeded to the next one, so make sure you have read everything thoroughly before continuing with the survey.

new page

Previous study - task

Participants in the previous study had to choose how many tasks they wanted to solve. A task consisted of counting frog emojis in a grid such as the one below. Participants could freely choose whether they wanted to solve no tasks, 1 task, or a maximum of 20 tasks.

What did a task look like?



Please count the frogs and enter the number here:

(This is only an example, please do not solve this task).

- Why would anyone choose to solve these tasks? As you can see, this task is rather annoying. In fact, 90% of participants found it annoying.
- However, for every real effort task they solved, we donated to **moorland protection** on their behalf. Every task they solved amounted to a donation of 5 eurocents to a local charity that works on restoring former moorland. Moorland conservation is a very effective natural climate action that reduces CO2 emissions. **The more tasks someone solves, the better for the climate.** 85% of participants thought that solving a task is a good contribution to environmental protection.

new page

Previous study - practice task

Later on, you will make predictions on how many of these tasks participants chose to solve. In order to make a sound prediction, please familiarize yourself with the task.

<grid of frogs and broccoli emojis with 20 lines>

Please count the frogs and enter their number here:

Tasks got harder with time: The first task was 20 lines long the twentieth task was 39 lines long. Guessing the answer was also of limited use because 5 seconds needed to pass between every submission. We made sure that participants had to actually count the emojis and could not use AI tools.

new page

Previous study – social judgment

Each participant had to choose how many out of 20 tasks they wanted to solve. Everyone had to make this choice twice under different circumstances. Next to climate protection (e.g. moorland conservation), there was another reason for participants to solve tasks. There was a high chance (2/3) that one of their **choices would be partially made public to all other participants.**

How was a choice made public?

- If a participant chose to solve x tasks or more, they would be announced to other participants as "protecting the climate".
- If a participant chose to solve less than x tasks, they would be announced to the other participants as "not protecting the climate".

The number x differs between the two situations. So depending on the situation, one had to do different amounts to be "protecting the climate". The exact values of x are given on the next screen.

The study took place on Zoom with students from the same university, some of whom may know each other. This is an illustration of how the public announcements looked like.

<graphic illustration of spotlighting and renaming>

The exact statement that was read aloud for each participant was either

"<Participant ID> is "protecting the climate" because he/she solves x tasks or more and thus protects x*5 dm^2 (square decimeters) or more of moorland for the climate."

or

"<Participant ID> is "not protecting the climate" because he/she solves less than x tasks and thus protects less than x*5 dm^2 (square decimeters) of moorland for the climate."

new page

Previous study – conditions

Participants had to make the same choice twice – under two different conditions. Let's call the two conditions condition A and condition B.

Condition A

- If a participant chose to solve **11 tasks or more**, they were announced to the other participants as "protecting the climate".
- If a participant chose to solve **less than 11 tasks**, they were announced to the other participants as "not protecting the climate".

Condition B

- If a participant chose to solve **17 tasks or more**, they were announced to the other participants as "protecting the climate".
- If a participant chose to solve **less than 17 tasks**, they were announced to the other participants as "not protecting the climate".

The only difference between the conditions is how many tasks you had to at least solve to be "protecting the climate".

Everyone made their respective choice under each condition and then one condition would be randomly selected with equal probability.* A randomly selected half of the participants had to make their decision in condition A first and then in condition B and the other half of the participants the other way around.

*In fact, participants only had to solve the number of tasks they committed to under the selected condition. They knew this in advance.

new page

Previous study – What could participants learn about others' choices?

- If condition A was selected, everyone learned only whether everyone else was solving more or less than 11 tasks and nothing else.
- If condition B was selected, everyone learned only whether everyone else was solving more or less than 17 tasks and nothing else.
- No one ever learned the exact number of tasks a participant chose to solve.
- Something was only ever announced about the decision in one situation - never about the decisions in both situations.

new page

Previous study - Meaning

Note that the meaning of the terms "protecting the climate" and "not protecting the climate" is different under the different conditions. For instance, "not protecting the climate" means solving **less than 11 tasks** under condition A, whereas it means solving **less than 17 tasks** under condition B.

new page

Previous study - Participants

All 255 participants in the study were students in Munich. Equal numbers of men and women took part. Strong norms prevailed on the topic of climate change: 90% of the participants said that, in general, they tried to protect the climate. Almost all participants said that others should protect the climate. Most participants felt that it played a role that others could partly see the decisions.

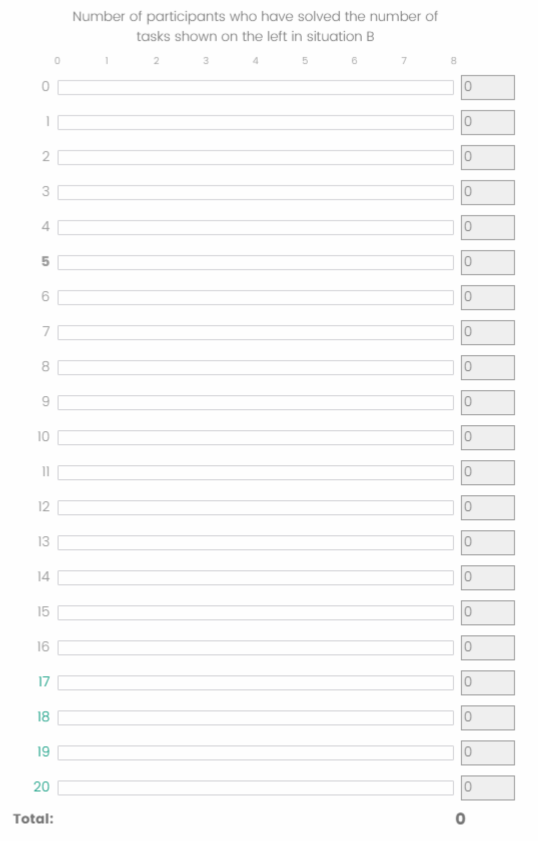
new page

What you will do

You are now ready to predict participants' choices. You will know participants' choices under condition A and will then be asked to predict participants' choices under condition B. For instance, there were 8 participants who chose to solve 5 tasks under condition A. What did these 8 people do under condition B? How many tasks did they choose under condition B?

- They were all "not protecting the climate" under condition A. Under condition B, they would be "protecting the climate" if they decided on 17 tasks or more, and they would be "not protecting the climate" if they decided on fewer than 17 tasks.
- The participants could choose exactly the same number, more or fewer tasks under condition B as under condition A.

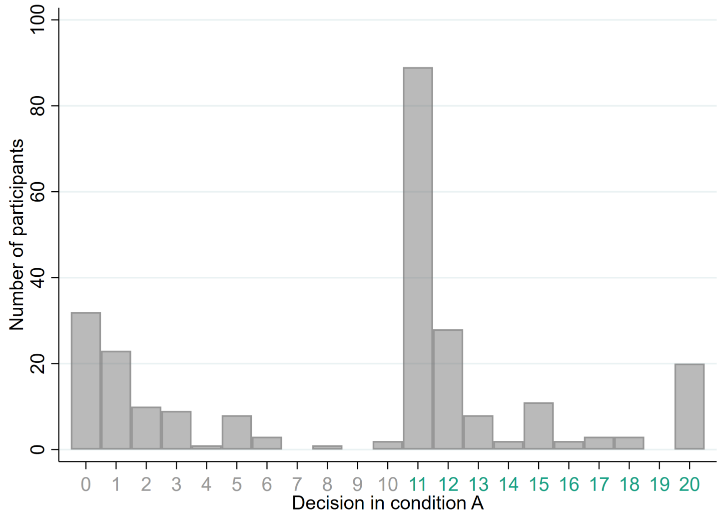
How many tasks did the 8 participants decide on under condition B?



new page

Previous study – Behavior under condition A

Here, you can see an overview of how participants behaved under condition A. For example, 89 participants chose 11 tasks under condition A - exactly the number of tasks that one had to solve at least to be "protecting the climate".



new page

Bonus

In addition to your 6 euros baseline payment for completing this questionnaire, you have a chance to win a 5 euros bonus payment based on the **accuracy of your predictions.** For each prediction, you will earn points. Specifically, the closer your prediction is to the actual behavior in the previous study, the more points you will earn.

At the end of this survey, we will sum all the points you earned and we will assign you the 5 euros bonus with a probability that depends on the sum total of your points. Specifically, **the more points you earned during this survey, the higher the probability that you will win the 5 euros bonus.**

The details of the point system used to determine your chance of winning the 5 Euro bonus are a bit complicated (they are explained below if you are interested). What is important to know is that the procedure by which the bonus is awarded ensures that it is in your best interest to make your prediction as accurate as you can.

The scoring system works as follows (in case you are interested). For each prediction, one of the bars is randomly selected. Based on your predicted value of this bar, it is calculated how many points you will receive for this prediction. This calculation follows the formula $y = \max\{100 - (100\frac{x}{n} - 100\frac{g}{n})^2, 0\}$, where x represents the true value, g represents the predicted value, and n represents the number of participants over whom the prediction is made. If the value you predict is the same as the true value, you get 100 points. The further it is from the true value, the fewer points you get. You will receive at least 0 points for each prediction. We will then average your points across all predictions and pay you the bonus with a probability equal to the average score. For example, if you have an average of 80 points across all predictions, you have an 80 out of 100 chance of winning the bonus.

new page

You will now answer some comprehension questions. Then you will make your predictions.

new page

Summary of previous study

- Tasks are annoying.
- The more tasks a participant chose to solve, the more money was donated to climate protection.
- Participants made a choice of how many tasks to solve twice (under condition A and condition B).
- Condition A: 11 tasks
  - If a participant chose to solve 11 tasks or more, they were announced to the other participants as "protecting the climate".
  - If a participant chose to solve less than 11 tasks, they were announced to the other participants as "not protecting the climate".
- Condition B: 17 tasks
  - If a participant chose to solve 17 tasks or more, they were announced to the other participants as "protecting the climate".
  - If a participant chose to solve less than 17 tasks, they were announced to the other participants as "not protecting the climate".
- If condition A was selected, everyone learned only whether everyone else was solving more or less than 11 tasks and nothing else.
- If condition B was selected, everyone learned only whether everyone else was solving more or less than 17 tasks and nothing else.
- No one ever learned the exact number of tasks a participant chose to solve.
- Only information about choices in one condition was made public – never about both conditions.

This summary was available to participants while answering the comprehension questions. This option will be represented by the button:

☐ Display summary of previous study

new page

Comprehension question 1

The more accurate your predictions, the higher your chances of winning the 5 euros bonus.

Reply options:

- *True*
- *False*

☐ Display summary of the previous study

new page

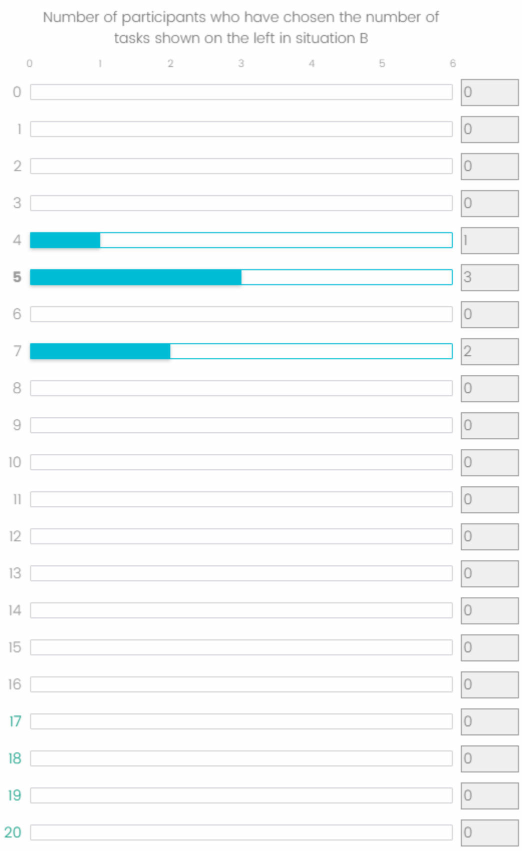
Comprehension question 2

Suppose you are asked about the participants who have chosen 5 tasks under condition A.

- Assume that 6 participants have chosen 5 tasks under condition A.
- They were all "not protecting the climate" under condition A.
- Under condition B, they would be:
  - "protecting the climate" if they chose 17 tasks or more
  - "not protecting the climate" if they chose less than 17 tasks
- The participants could choose exactly the same number, more or fewer tasks under condition B than under condition B.

How would you express that, of the 6 participants,

- 1 participant chose 4 tasks under condition B,
- 3 participants again chose 5 tasks under condition B,
- and 2 participants chose 7 tasks under condition B?



new page

Comprehension questions 3 and 4

Imagine a participant who does not like solving tasks but wants to be announced to others as "protecting the climate".

What is the minimum amount of tasks this participant has to solve in order to be "protecting the climate" under condition A?

What is the minimum amount of tasks this participant has to solve in order to be "protecting the climate" under condition B?

new page

Comprehension question 5

No one ever learned anything about both choices of a participant.

Reply options:

- *True*
- *False*

——— *new page* ———

Comprehension question 6

Which statement is correct?

Reply options:

- *Other participants only learned whether a participant chose to solve more or less than x tasks.*
- *Other participants learned exactly how many tasks every participant chose to solve.*

——— *new page* ———

Thank you. You have now answered all comprehension questions correctly. Please make your predictions about how participants behaved under condition B, given how they behaved under condition A, on the following screens.

——— *new page* ———

II Predictions

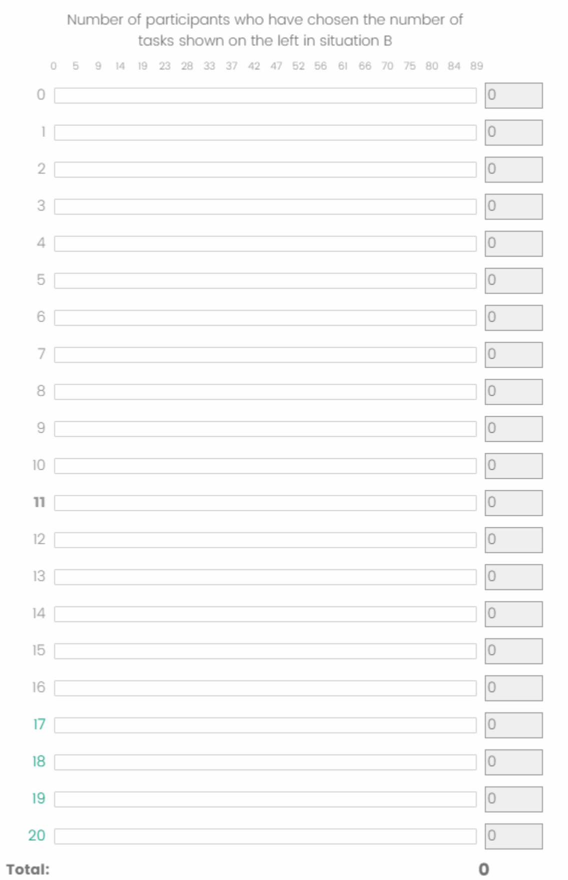
Each subject was presented with 21 screens similar to the one below. They had to make predictions on what those participants who chose 0,1,2,...,20 tasks under condition A chose under condition B. Below is an example screen for the Predict Broad Convention condition and the Predict Narrow Convention condition.

——— *predict broad convention* ———

Participants who chose 11 tasks under condition A

- 89 participants choose 11 tasks under condition A.
- They were all "protecting the climate" under condition A.
- Under condition B they would be "protecting the climate" if they chose 17 tasks or more.
- In condition B the participants could decide on exactly the same number, more or fewer tasks than under condition A.

How many tasks did the 89 participants choose to protect under condition B?

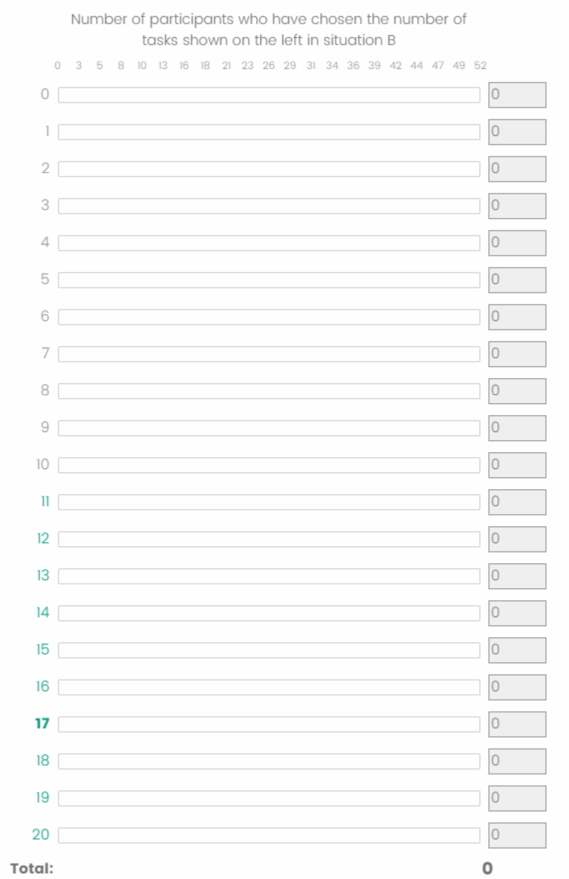


predict narrow convention

Participants who chose 17 tasks under condition A

- 52 participants chose 17 tasks under condition A.
- They were all "protecting the climate" under condition A.
- Under condition B they would be "protecting the climate" if they chose 11 tasks or more.
- In condition B the participants could decide on exactly the same number, more or fewer tasks than under condition A.

How many tasks did the 52 participants choose to protect under condition B?



new page

III Natural Action

Thank you very much. You have made all your predictions. We now ask you to choose for yourself how many tasks you want to solve. No one could (even partially) observe your decision. You simply choose how many tasks you want to solve.

new page

What can you choose?

When you choose, you have the opportunity to protect up to one square meter for the rewetting of moorlands. You can choose how many of a maximum of 20 tasks you would like to solve. For each task solved, 5 eurocents will be donated to BUND Naturschutz e.V.. Every euro donated enables the purchase of one square meter of moorland for rewetting and thus for **climate protection (1 task solved: 5dm^2 (square decimeters) of moorland).** By solving one task, approximately as much CO2 can be stored in the earth as is emitted when driving 3 km by car.

- 1 task solved: 5 dm$^{\mathbf{2}}$ of protected bog area.

- With a probability of 33% (one third), you have to solve the number of tasks you now set. Only then will a donation be made.
- The tasks become increasingly difficult. The first task is 20 lines long (as below), the second task is 21 lines long, ..., and the twentieth task is 39 lines long.
- The fewer tasks you solve, the sooner you finish the survey. Without any tasks, the rest of the survey will take about 5 minutes.
- Nobody can observe your decision (even partially).

You have already solved a practice task. As a reminder, here is what a task looks like.
<grid of frogs and broccoli emojis>
Please count the frogs and enter their number here:
(This is only an example, please do not solve this task.)

new page

Your decision
How many tasks do you want to solve?
<slider from 0 to 20>

- 1 task solved:5 dm² of protected bog area.
- With a probability of 33% (one third), you have to solve the number of tasks you now set. Only then will you also make a donation.
- The tasks will become increasingly difficult.
- The fewer tasks you solve, the sooner you are finished with the survey.
- Nobody can observe your decision (even partially).

IV Real Effort Tasks

(with probability $\frac{1}{3}$)

It has been randomly chosen that you have to solve tasks. These will be followed by a short questionnaire and then you will be redirected to the page for payment.
Between 0 and 20 such screens, depending on the subject's choice for the randomly selected situation.
<grid of frogs and broccoli emojis>
Please count the frogs and enter their number here:

——— *or (with probability $\frac{2}{3}$)* ———

It has been randomly chosen so that you do not have to solve any tasks.

V Questionnaire on Norms and Demographics

Thank you very much. We now kindly ask you to answer a few more questions. You will then be redirected to the payment page.

——— *new page* ———

Do you try to protect the climate?
Reply options:

- *Yes*
- *No*

Do you think students in Munich should try to protect the climate?
Reply options:

- *Yes*
- *No*

——— *new page* ———

We are asking the same two questions to many students in Munich. What do you think?

How many of the students in Munich say they try to protect the climate?
<slider 0,1...,100(%)>

How many of the students in Munich say that students in Munich should try to protect the climate?
<slider 0,1...,100(%)>

——— *new page* ———

How much do you care about the climate?
Reply options:

- *Very little.*
- *Rather little.*
- *Rather much.*
- *Very much.*

new page

Imagine that you are in a competition with 50 randomly selected students in Munich. You all have to enter a number between 0 and 100. The average is calculated from your entries. The number closest to 2/3 of the average wins the competition (in case of a tie, the competition is protected by drawing lots).
What number would you enter?

new page

Which gender do you identify with?
Reply options:

- *female*
- *male*
- *non-binary*
- *I don't know*

new page

Are you a university student?
Reply options:

- *Yes*
- *No*

new page

Do you agree with the following statement? This survey was easy to understand.
Reply options:

- *Strongly disagree*
- *Disagree*
- *Agree*
- *Strongly agree*

new page

If there is anything else you would like to tell us (if you had difficulty understanding the questions, if you had problems with the content or format of the survey, etc.), please do so here:

Appendix to Chapter 2

B.1 Mathematical Appendix

Proof of Proposition 1

Given a sender's disclosure strategy as fully described by $\mu = P(m = \emptyset|\omega = 1)$, the receivers' posterior upon seeing the empty message is $q(m = \emptyset) = P(\omega = 1|m = \emptyset) = \frac{P(m=\emptyset|\omega=1)P(\omega=1)}{P(m=\emptyset)}$. Trivially, $q(m = h) = P(\omega = 1|m = h) = 1$. This yields the sender's expected payoffs of

$$\begin{aligned}\mathbb{E}[U_S(\mu, q)] = p((1-\mu)\alpha + \mu(\alpha q(m = \emptyset) + (1-\alpha)(1 - q(m = \emptyset)))) \\ +(1-p)(1 - q(m = \emptyset))\end{aligned}$$

Let's first show that in equilibrium $\mu \notin (0, 1)$. We want to show that for every $\mu \in (0, 1)$ and given the induced posterior beliefs for this μ, it is profitable for the sender to change her disclosure strategy. Since the receivers' posterior beliefs q will not change in response to the deviation, we fix the family of posterior beliefs at $\bar{q}$.

$\alpha > \frac{1}{2} \iff \frac{\partial \mathbb{E}[U_S(\mu,\bar{q})]}{\partial \mu} < 0$, meaning that it is beneficial for the sender to deviate by increasing her disclosure frequency $(1 - \mu) = P(m = h|\omega = 1)$.

$\alpha < \frac{1}{2} \iff \frac{\partial \mathbb{E}[U_S(\mu,\bar{q})]}{\partial \mu} > 0$, meaning that it is beneficial for the sender to deviate by decreasing her disclosure frequency $(1 - \mu) = P(m = h|\omega = 1)$.

Iff $\alpha = \frac{1}{2}$, the sender is indifferent between any value of μ, in which case we assume her to be truthful (e.g. $\mu^* = 0$).

Note that the above implies that there are no profitable deviations for $\mu^* = 0$ if $\alpha \geq \frac{1}{2}$ and for $\mu^* = 1$ if $\alpha < \frac{1}{2}$. $\square$

Proof of Proposition 2

In the presence of co-audience neglect, the sender's disclosure strategy μ, as fully described by $\mu = P(m = \emptyset|\omega = 1)$, does not affect receivers' posteriors. Receivers' posteriors are given by $q_{i\in\{a,n\}}(m = h) = 1$, $q_a(m = \emptyset) = 0$ and $q_n(m = \emptyset) = p = P(\omega = 1)$. This yields the sender's expected payoffs of

$$\mathbb{E}[U_S(\mu, q)] = p((1-\mu)\alpha + \mu(1-\alpha)(1-p)) + (1-p)(\alpha + (1-\alpha)(1-p))$$

where $\mathbb{E}[U_S(\mu = 0, q)] = \alpha p + (1-p)(\alpha + (1-\alpha)(1-p))$ and $\mathbb{E}[U_S(\mu = 1, q)] = p(1-\alpha)(1-p) + (1-p)(\alpha + (1-\alpha)(1-p))$. Like in the previous proof, we first show that in equilibrium $\mu \notin (0, 1)$.
$\alpha > \frac{1-p}{2-p} \iff \frac{\partial \mathbb{E}[U_S(\mu,q)]}{\partial \mu} < 0$, meaning that it is beneficial for the sender to deviate by increasing her disclosure frequency $(1-\mu) = P(m = h|\omega = 1)$.
$\alpha < \frac{1-p}{2-p} \iff \frac{\partial \mathbb{E}[U_S(\mu,q)]}{\partial \mu} > 0$, meaning that it is beneficial for the sender to deviate by decreasing her disclosure frequency $(1-\mu) = P(m = h|\omega = 1)$.
Iff $\alpha = \frac{1-p}{2-p}$, the sender is indifferent between any value of μ, in which case we assume her to be truthful (e.g. $\mu^{**} = 0$).
Note that the above implies that there are no profitable deviations for $\mu^{**} = 0$ if $\alpha \geq \frac{1-p}{2-p}$ and for $\mu^{**} = 1$ if $\alpha < \frac{1-p}{2-p}$. □

Partial Co-audience Neglect

We denote the share of co-audience-neglecting receivers by $\gamma \in [0, 1]$. Their beliefs are $q_{i\in\{a,n\}}(m = h) = 1$, $q_a(m = \emptyset) = 0$ and $q_n(m = \emptyset) = p = P(\omega = 1)$. Rational receivers' beliefs are Bayes-rational given the sender's disclosure strategy $\mu = P(m = \emptyset|\omega = 1)$ inducing a posterior of $q(m = \emptyset) = P(\omega = 1|m = \emptyset) = \frac{P(m=\emptyset|\omega=1)P(\omega=1)}{P(m=\emptyset)}$ that we denote by $\hat{q}$ in what follows. Rational receivers are assumed to have a correct perception of the sender's incentives (like the sender, they have correct higher-order beliefs) and of the share of co-audience neglecting receivers. We also assume that the two random events leading to a receiver's type determination, governed by α and γ, are independent. The sender's expected payoffs are thus

$$\begin{aligned}\mathbb{E}[U_S(\mu, \hat{q})] = p((1-\mu)\alpha + \mu(\alpha(1-\gamma)\hat{q} + (1-\alpha)(1-\gamma p + (\gamma-1)\hat{q}))) + \\ (1-p)(\alpha(1 + (\gamma-1)\hat{q}) + (1-\alpha)(1-\gamma p + (\gamma-1)\hat{q}))\end{aligned}$$

Again, we show that in equilibrium $\mu \notin (0,1)$. We do so by showing that for every $\mu \in (0,1)$ and given the receivers' posterior beliefs, which are partially induced by μ, it is profitable for the sender to change μ. As in the proof of Proposition 1, we can fix the family of receivers' beliefs at $\bar{q}$, the value that corresponds to the candidate value of μ.
$\alpha > \frac{1-\gamma p-(1-\gamma)\hat{q}}{2-\gamma p-2(1-\gamma)\hat{q}} \iff \frac{\partial \mathbb{E}[U_S(\mu,\bar{q})]}{\partial \mu} < 0$, meaning that it is beneficial for the sender to deviate by increasing her disclosure frequency $(1-\mu) = P(m=h|\omega=1)$.
$\alpha < \frac{1-\gamma p-(1-\gamma)\hat{q}}{2-\gamma p-2(1-\gamma)\hat{q}} \iff \frac{\partial \mathbb{E}[U_S(\mu,\bar{q})]}{\partial \mu} > 0$, meaning that it is beneficial for the sender to deviate by decreasing her disclosure frequency $(1-\mu) = P(m=h|\omega=1)$.
Iff $\alpha = \frac{1-\gamma p-(1-\gamma)\hat{q}}{2-\gamma p-2(1-\gamma)\hat{q}}$, the sender is indifferent between any value of μ in which case we assumed her to be truthful ($\mu^* = 0$).
Letting $\kappa \equiv \frac{1-\gamma p-(1-\gamma)\hat{q}}{2-\gamma p-2(1-\gamma)\hat{q}}$, we note that κ reduces to $\frac{1}{2}$ for $\gamma = 0$ (BNE) and to $\frac{1-p}{2-p}$ for $\gamma = 1$ (CANE).
The above implies that there are no profitable deviations for $\mu^* = 0$ if $\alpha \geq \kappa$ and for $\mu^* = 1$ if $\alpha < \kappa$. Also, $\frac{\partial \kappa}{\partial \gamma} < 0$ implies that for intermediate values of γ (i.e. $\gamma \in (0,1)$), the sender's equilibrium threshold for being truthful will be between the threshold in the BNE and CANE. This means that for our experimental parameterization (see Figure B.1), the qualitative predictions for the sender disclosure strategy of BNE versus CANE coincide with those of BNE versus partial CANE. □

B.2 Tables

Table B.1: Session Overview

Session	Date	Started with	Screened out	Not matched into groups	Timeouts during experiment	Completed
1	March 2, 2021	42	19	8	0	15
2	March 2, 2021	39	11	1	3	24
3	March 4, 2021	21	7	2	0	12
4	April 14, 2021	53	15	7	10	21
5	April 22, 2021	57	3	3	3	48
6	April 22, 2021	55	4	3	0	48
7	April 23, 2021	53	3	2	3	45
8	April 23, 2021	54	4	2	3	45
9	April 23, 2021	56	4	4	0	48
		430	70	32	22	306

Notes: From session 5 onward, we started to allow for one mistake in the comprehension questions. Participants who timed out during the main part of the experiment are included in the analysis, as pre-registered. For our main sample $n = 328$, and for our pre-registered sample $n = 280$.

Table B.2: Balance

Variable	(1) Aligned receiver N	Mean/SE	(2) Non-aligned receiver N	Mean/SE	(3) Sender N	Mean/SE	T-test Difference (1)-(2)	(1)-(3)	(2)-(3)
Age	102	23.569 (0.525)	102	23.922 (0.468)	102	23.441 (0.325)	-0.353	0.127	0.480
Female	110	0.536 (0.048)	110	0.645 (0.046)	108	0.509 (0.048)	-0.109	0.027	0.136**
Political affiliation	102	4.775 (0.180)	102	4.657 (0.168)	102	5.098 (0.182)	0.118	-0.324	-0.441*
Social media usage	102	2.559 (0.121)	102	2.578 (0.118)	102	2.539 (0.120)	-0.020	0.020	0.039
Participation record	102	2.794 (0.106)	102	2.657 (0.114)	102	2.843 (0.113)	0.137	-0.049	-0.186
High school GPA	102	2.046 (0.063)	102	1.965 (0.061)	102	1.980 (0.063)	0.081	0.066	-0.016
High school math grade	102	2.278 (0.095)	102	2.206 (0.086)	102	2.188 (0.087)	0.073	0.090	0.018
Net income	102	1.608 (0.103)	102	1.363 (0.071)	102	1.441 (0.099)	0.245*	0.167	-0.078
Minority	102	0.765 (0.042)	102	0.775 (0.042)	102	0.814 (0.039)	-0.010	-0.049	-0.039

Notes: The value displayed for t-tests are the differences in the means across the groups. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table B.3: Balance - Pre-registered Sample Size

Variable	(1) Aligned receiver N	Mean/SE	(2) Non-aligned receiver N	Mean/SE	(3) Sender N	Mean/SE	T-test Difference (1)-(2)	(1)-(3)	(2)-(3)
Age	86	23.628 (0.611)	86	24.070 (0.535)	86	23.116 (0.348)	-0.442	0.512	0.953
Female	94	0.521 (0.052)	94	0.617 (0.050)	92	0.511 (0.052)	-0.096	0.010	0.106
Political affiliation	86	4.733 (0.189)	86	4.709 (0.191)	86	5.093 (0.198)	0.023	-0.360	-0.384
Social media usage	86	2.558 (0.132)	86	2.674 (0.133)	86	2.581 (0.131)	-0.116	-0.023	0.093
Participation record	86	2.849 (0.116)	86	2.651 (0.128)	86	2.860 (0.121)	0.198	-0.012	-0.209
High school GPA	86	2.005 (0.064)	86	1.930 (0.065)	86	1.976 (0.069)	0.074	0.029	-0.045
High school math grade	86	2.252 (0.105)	86	2.223 (0.096)	86	2.228 (0.095)	0.029	0.024	-0.005
Net income	86	1.547 (0.105)	86	1.384 (0.074)	86	1.372 (0.103)	0.163	0.174	0.012
Minority	86	0.744 (0.047)	86	0.791 (0.044)	86	0.837 (0.040)	-0.047	-0.093	-0.047

Notes: The value displayed for t-tests are the differences in the means across the groups. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table B.4: Receiver Beliefs - Pre-registered Sample Size

	(1) Belief	(2) Belief	(3) Belief	(4) Belief	(5) Belief*	(6) Belief*
Aligned receiver (d)	-4.016	-3.610	-4.273	-4.012	-4.181	-4.065
	(2.615)	(3.217)	(3.002)	(2.504)	(2.539)	(2.762)
High alignment (d)		-25.12***		-22.99***	-27.53***	
		(3.504)		(2.893)	(2.619)	
Aligned receiver x High alignment		-0.816				
		(5.045)				
Rounds 5 to 8 (d)			-3.253	-0.411		
			(2.333)	(1.956)		
Aligned receiver x Rounds 5 to 8			0.526			
			(3.219)			
High alignment x Rounds 5 to 8				-5.185		
				(3.182)		
High state at t-1 (d)					4.241**	
					(2.049)	
High state and disclosure at t-1 (d)					-2.519	
					(2.175)	
No disclosure experience (d)						1.054
						(2.750)
Aligned receiver x No disclosure exp.						0.235
						(3.760)
Constant	42.46***	54.97***	44.06***	55.37***	54.58***	42.20***
	(1.854)	(2.351)	(2.211)	(2.282)	(2.233)	(1.893)
Observations	1420	1420	1420	1420	1232	1420
R^2	0.004	0.172	0.006	0.176	0.199	0.004

Notes: Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors are clustered at the participant level. One observation is one receiver belief in the high state upon the empty message, i.e. $q_i = P(\omega = 1|m = \emptyset)$. An asterisk next to an outcome variable indicates that the regression model was not pre-registered.

Table B.5: Heterogeneity in Receiver Beliefs - Pre-registered Sample Size

	(1) Belief rounds > 1*	(2) Belief rounds > 1*	(3) Belief*
Aligned receiver (d)	-1.259 (2.949)	-0.0870 (3.024)	-9.859** (4.745)
CAN Type (5pp deviation) (d)	0.0761 (3.623)		
Aligned receiver x CAN Type (5pp)	-19.28*** (5.787)		
CAN Type (10pp deviation) (d)		0.130 (3.410)	
Aligned receiver x CAN Type (10pp)		-19.50*** (5.273)	
Extreme political views			-2.148 (1.459)
Aligned receiver x Extreme political views			3.590 (2.171)
Constant	42.25*** (2.181)	42.24*** (2.265)	45.76*** (3.467)
Observations	1232	1232	1376
R^2	0.028	0.036	0.009

Notes: Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors are clustered at the participant level. One observation is one receiver belief in the high state upon the empty message, i.e. $q_i = P(\omega = 1|m = \emptyset)$. An asterisk next to an outcome variable indicates that the regression model was not pre-registered.

Table B.6: Sender Beliefs and Strategy - Pre-registered Sample Size

	(1) Sender belief	(2) Disclosure	(3) Disclosure
High alignment (d)	-3.407	0.571***	0.565***
	(2.137)	(0.0519)	(0.0673)
W.r.t. aligned receiver (d)	-15.26***		
	(2.081)		
Rounds 5 to 8 (d)	-4.203		-0.0193
	(3.410)		(0.0686)
W.r.t. aligned receiver x Rounds 5 to 8	0.162		
	(1.683)		
High alignment x Rounds 5 to 8			0.0116
			(0.0804)
Constant	24.90***	0.294***	0.303***
	(3.689)	(0.0393)	(0.0489)
Observations	1406	354	354
R^2	0.083	0.334	0.334

Notes: Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors are clustered at the participant level. In column (1), one observation corresponds to one sender belief about one receiver belief, specifically her belief about $q_i = P(\omega = 1|m = \emptyset)$. In columns (2) and (3), one observation corresponds to one sender disclosure decision. Remember, a sender had to make a disclosure decision only when the state was high.

Table B.7: Robustness

	(1) Belief	(2) Belief	(3) Belief	(4) Belief	(5) Belief	(6) Belief	(7) Belief	(8) Belief
Aligned receiver (d)	-4.255*	-3.718**	-4.597*	-4.151*	-5.036*	-5.534**	-4.152*	-4.537*
	(2.554)	(1.810)	(2.646)	(2.473)	(2.651)	(2.565)	(2.434)	(2.193)
Net income	0.514							
	(1.411)							
Female (d)							3.401	
							(2.435)	
Constant	42.29***	23.06***	42.89***	42.98***	43.97***	43.46***	41.10***	43.40***
	(2.901)	(1.303)	(1.925)	(1.790)	(1.888)	(1.884)	(2.306)	(1.248)
Observations	1568	1071	1353	1632	1403	1501	1676	1676
R^2	0.004	0.011	0.005	0.004	0.006	0.008	0.008	0.005

Standard errors in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Notes: Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Column (1) controls for net income, which was not balanced among receiver types (see Table B.2). Column (2) excludes all clearly irrational beliefs larger than 50%. Column (3) excludes all participants who made one mistake in the control questions (a maximum of one mistake was allowed). Column (4) restricts the sample to those who completed the entire experiment (204 receivers). Column (5) excludes small sessions with fewer than 24 participants. Column (6) excludes those who clicked fastest through the instructions (fastest 10%). Column (7) adds a female dummy as a standard control dummy which was almost unbalanced across treatment. Column (8) clusters standard errors on the session level. In all other columns, standard errors are clustered on the participant level.

Table B.8: Type Classification Robustness

	(1) Belief rounds > 1*	(2) Belief rounds > 1*	(3) Belief rounds > 1*	(4) Belief rounds > 1*
Aligned receiver (d)	-2.332	-1.404	-2.087	-1.287
	(2.712)	(2.772)	(2.697)	(2.739)
CAN Type (5pp deviation) (d)	-0.531			
	(3.913)			
Aligned receiver x CAN Type (5pp)	-15.12**			
	(6.387)			
CAN Type (10pp deviation) (d)		-0.533		
		(3.610)		
Aligned receiver x CAN Type (10pp)		-15.98***		
		(5.627)		
CAN Type (6pp deviation symmetric) (d)			1.008	
			(4.048)	
Aligned receiver x CAN Type (6pp symmetric)			-16.66**	
			(6.471)	
CAN Type (10pp deviation symmetric) (d)				0.0759
				(3.802)
Aligned receiver x CAN Type (10pp symmetric)				-16.59***
				(5.753)
Constant	43.01***	43.03***	42.76***	42.91***
	(2.020)	(2.085)	(1.999)	(2.041)
Observations	1456	1456	1456	1456
R^2	0.021	0.027	0.021	0.027

Notes: Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors are clustered at the participant level. One observation is one receiver belief in the high state upon the empty message (i.e. $q_i = P(\omega = 1|m = \emptyset)$). An asterisk next to an outcome variable indicates that the regression model was not pre-registered. Column (1) defines a CAN type as an aligned receiver that stated a belief $q_a \in [0, 5]$ in round 1 or a non-aligned receiver that stated a belief $q_n \in [45, 50]$ in round 1. Column (2) defines a CAN type as an aligned receiver that stated a belief $q_a \in [0, 10]$ in round 1 or a non-aligned receiver that stated a belief $q_n \in [40, 50]$ in round 1. Column (3) defines a CAN type as an aligned receiver that stated a belief $q_a \in [0, 6]$ in round 1 or a non-aligned receiver that stated a belief $q_n \in [47, 53]$ in round 1. Column (4) defines a CAN type as an aligned receiver that stated a belief $q_a \in [0, 10]$ in round 1 or a non-aligned receiver that stated a belief $q_n \in [40, 50]$ in round 1.

B.3 Figures

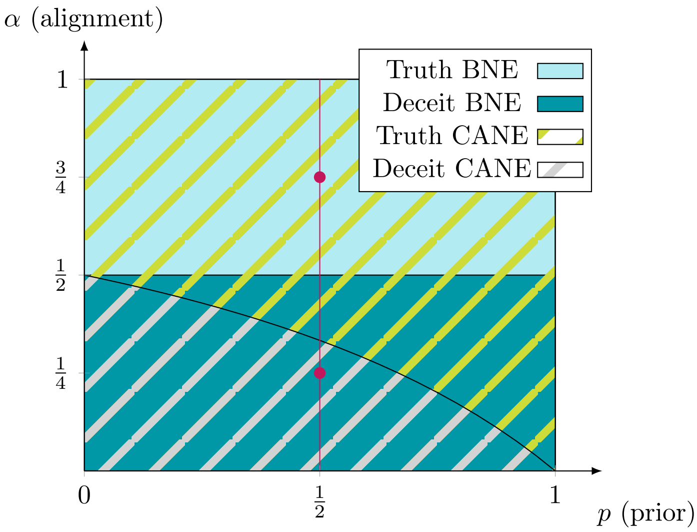


Figure B.1: Sender's Disclosure Strategy

Notes: BNE stands for Bayesian Nash Equilibrium and CANE stands for Co-audience Neglect Equilibrium. The red circles identify our experimental parameterization, i.e. the tuples of prior and average alignment (α, p) in the two alignment conditions, $(\frac{1}{2}, \frac{3}{4})$ in the high alignment condition and $(\frac{1}{2}, \frac{1}{4})$ in the low alignment condition.

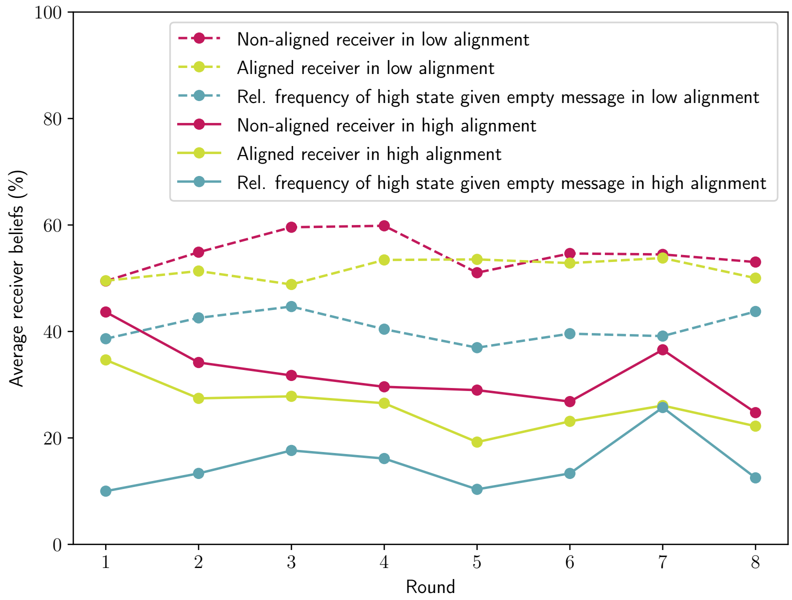


Figure B.2: Average Receiver Beliefs Across Rounds

Notes: Average receiver beliefs upon the empty message compared to the relative frequency of the high state given the empty message.

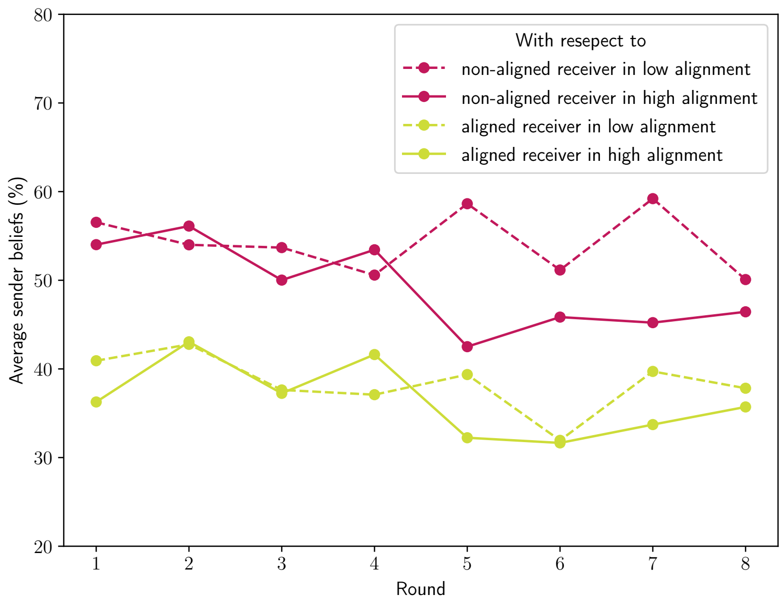


Figure B.3: Sender Beliefs

Notes: Sender beliefs about receiver beliefs, i.e. sender beliefs on $q_{i\in\{a,n\}} = P(\omega = 1|m = \emptyset)$.

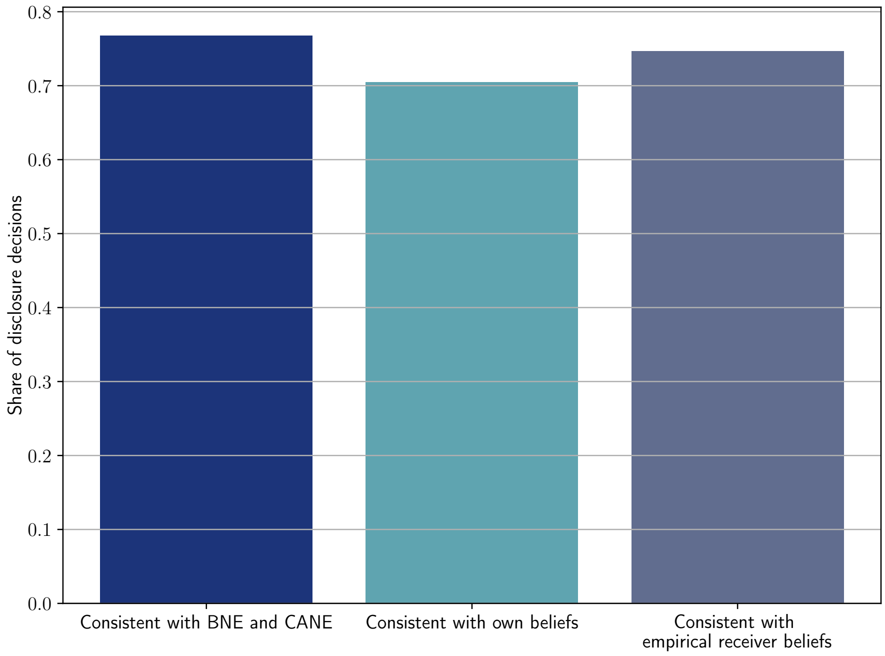


Figure B.4: Sender's Disclosure Strategy

Notes: We observe 430 disclosure decisions. A disclosure decision is consistent with BNE and CANE if the verifiable message is chosen in the high alignment condition and the empty message is chosen in the low alignment condition. 77% of sender disclosure decisions are consistent with BNE and CANE. A disclosure decision is consistent with a sender's own beliefs, if the disclosure decision is payoff maximizing given her stated beliefs on receiver beliefs. 70% of sender disclosure decisions are consistent with own beliefs. A disclosure decision is consistent with empirical beliefs if the disclosure decision is payoff-maximizing given the average receiver beliefs in that round, pooled across sessions. 75% of sender disclosure decisions are consistent with empirical beliefs.

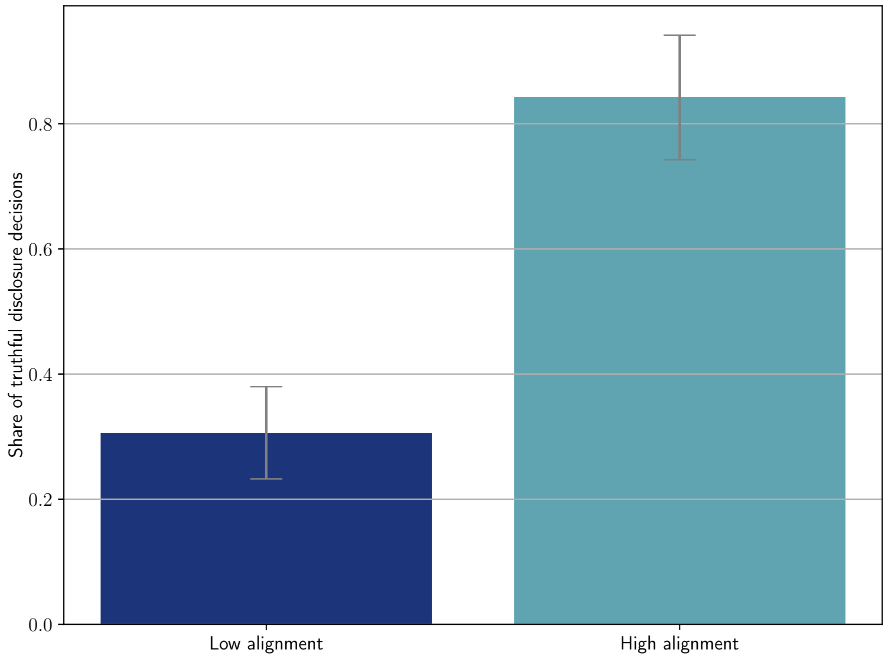


Figure B.5: Sender Disclosure by Alignment

Notes: In the low (high) alignment condition 31% (84%) of disclosure decisions are truthful. The error bars depict 95% confidence intervals.

B.4 Experimental Instructions

I Attention checks

Welcome!

When you click Next, the first stage begins. Only if you correctly complete the following short attention tests can you participate in the actual experiment. If you fail an attention test, you will be excluded from all payments. On each page, you will see a timer at the top. Make your entries and press Continue before the timer runs out. If you fail, you will be excluded from the experiment. (If you are excluded, you can simply close the browser window).

new page

Attention check 1/3

Please type the word you see written above into the text box.

new page

Attention check 2/3

<pictures of fruits with numbers on them>

Please enter the number that is written on the avocado.

new page

Attention check 3/3

Three neighbors live along a street in three colorful houses. Mrs. Green lives in a yellow house, Mr. Black lives in a blue house and Mrs. Purple lives in a green house. What color is Mrs. Green's house?

Reply options:

- *yellow*
- *blue*
- *red*
- *green*

new page

You have successfully completed the attention tests.

Congratulations, you have successfully completed all the attention tests. Click Next so that you can start the actual experiment.

new page

II Instructions

Welcome to this experiment!

Thank you for your participation! We ask you to participate in this experiment without interruption, not to take breaks, and not to communicate with others during the experiment. You should allow approximately 60 minutes for the duration of the experiment.

On each page, you will see a timer at the top. Make your entries and press Continue before the timer runs out. Also, the first part of the experiment should take you no more than 20 minutes. This is a generous amount of time. If you do not make it, you will be excluded from the experiment. (If you are excluded, you can simply close the browser window.)

During the experiment, you and all other participants will be asked to make choices. Both your own decisions and those of the participants connected to you will determine your payoff. For completing a questionnaire at the end of the experiment, you will receive an additional 2 euros. Your payment will be paid to you via your chosen channel following the experiment. However, you will only be paid out if you have completed the experiment in full.

All your decisions and answers remain anonymous. You won't know who you interact with or how much other participants earn. Once you click Continue, the instructions will follow. Take your time to understand them thoroughly. Afterward, you will be asked control questions about the instructions. Only if you answer all of them correctly can you continue with the experiment.

The following is on every screen.

We only use male designations. These are to be understood as gender-neutral. In case of technical problems or content-related questions, please contact <email address>.

——— *new page* ———

Random role assignment. Example for partner (aligned receiver).

Your role

In this experiment, you interact in a group of three with two other randomly selected participants. The experiment consists of 8 rounds. In each round, you will be randomly assigned new participants to interact with. It is very unlikely that you will interact with the same participants in successive rounds. Each participant in your group of three has a role. There is a gold miner, a partner, and a bandit.

You were assigned the role of partner.

You remain in this role for the entire experiment. Except for this notification of your role, the instructions are the same for all participants in terms of content.

——— *new page* ———

The interaction

The interaction with the other participants in your group is about whether the gold miner in your group finds gold or does not find gold.

The gold miner either finds gold or does not find gold. Both states have a 50 percent probability of occurring. In each round, it is decided anew whether the gold miner finds gold or no gold.

☐ Show role and previous instructions

——— *new page* ———

The gold miner's action

When the gold miner finds gold, he must decide whether to tell the partner and the bandit about him finding gold. The gold miner can then choose between two messages: "gold" and "no gold".

The chosen message is always sent to both - the partner and the bandit. The gold miner cannot send different messages to the partner and the bandit.

<graphic illustration of game tree>

The partner and the bandit see the same message from the gold miner. They either see the message “no gold” if the gold miner has not found any gold or he has decided not to share that he found gold. Or they see the message “gold” if the gold miner has found gold and has decided to share that he found gold.

☐ Show role and previous instructions

new page

The actions of the partner and the bandit

Before the partner and bandit see the actual message from the gold miner, they must indicate how likely they would think it is that the gold miner has found gold if they received the “no gold” message from the gold miner.

This guess ranges from 0 to 100 percent.

It may be that the gold miner actually found no gold. However, it is also possible that he has found gold but has decided not to report this.

If the gold miner sends the message “gold”, he must have actually found gold. Thus, the probability of finding gold after this message is 100%. This guess is already logged in and cannot be changed by the partner and bandit.

After that, the partner and bandit see which message the gold miner actually sent them and then get paid for the guess they stated for that message.

☐ Show role and previous instructions

new page

Example for bandit (non-aligned receiver).

Goals of participants

The participants have different goals depending on what role they are in.

The gold miner wants to convince you, the bandit, that he has not found any gold - regardless of whether he has actually found gold or not.

At the same time, he wants to tell the partner the truth, namely whether he has found gold or not. However, he can only send **one and the same message** to both of them together - you and the partner.

In some rounds, it is more important to the gold miner what you - the bandit - believe and in some rounds, it is more important to the gold miner what the partner believes.

You want to find out whether the gold miner has found gold or not. The partner also wants to guess as correctly as possible whether the gold miner has found gold or not.

<graphic illustration of incentives>

These goals result from how the participants are paid. This is described in detail on the next screen.

☐ Show role and previous instructions

new page

Payment

You can earn lottery tickets in every round. The more lottery tickets you earn, the more money you earn on average.

With your lottery tickets you can win a prize of 8 euros. You start each round with an urn containing 400 rivets. During a round, these rivets can be replaced with lottery tickets. At the end of the experiment, a round is randomly selected and a ticket is drawn from your urn from that round. Only if a lottery ticket is drawn you will receive the prize of 8 euros. So the more lottery tickets you earn, the higher your chance of winning the prize of 8 euros.

new page

Payment of the gold miner

The gold miner is paid depending on the guesses of the partner and the bandit. In the following, we simply call the probability of finding gold given by the bandit "guess_bandit". So the bandit believes that the gold miner has found gold with a probability of "guess_bandit". We call the guess given by the partner "guess_partner". Both are numbers between 0 and 100.

The payment of the gold miner depends on the bandit:

The gold miner gets (100 - guess_bandit) lottery tickets, regardless of whether he found gold or not. So the less the bandit believes that the gold miner has found gold, the more the gold miner earns.

The payment of the gold miner also depends on the partner:

The gold miner gets additional guess_partner lottery tickets if he finds gold and (100 - guess_partner) lottery tickets if he doesn't find gold. This means that the closer the partner's guess is to the truth, the more the gold miner earns.

From round to round, it is randomly decided which of the two guesses - the bandit's or the partner's - is more important for the gold miner:

With probability 50%, the bandit's guess is more important for the gold miner. Then the part of the gold miner's payment that relates to the bandit is **tripled**. With probability 50%, the partner's guess is more important for the gold miner. Then the part of the gold miner's payment that relates to the partner is **tripled**.

If the bandit's guess is more important for the gold miner: Total payment of the gold miner in lottery tickets =

3 x (100 - guess_bandit) + (guess_partner), if he finds gold

3 x (100 - guess_bandit) + (100 - guess_partner), if does not find gold

If the partner's guess is more important for the gold miner: Total payment of the gold miner in lottery tickets =

(100 - guess_bandit) + 3 x (guess_partner), if he finds gold

(100 - guess_bandit) + 3 x (100 - guess_partner), if does not find gold

new page

Payment of the partner and the bandit

The partner and bandit are paid according to the accuracy of their guesses. Each of them gets more lottery tickets the closer his guess is to the truth.

Total payment of the partner and bandit in lottery tickets =

$400 - 4/100(guess - 100)^2$, if gold miner finds gold

$400 - 4/100(guess)^2$, if gold miner does not find gold

This formula may look a bit complicated. The only important thing for you to know is that it is worthwhile for the partner and the bandit to state their true guess. They maximize their probability of winning the prize of 8 euros if they state the probability they are really convinced of.

☐ Show role and previous instructions

new page

As soon as a participant answers one comprehension question incorrectly, the participant is excluded and gets the show-up fee of 6 euros. From session 5 onwards, we allowed for one mistake. Example for bandit (non-aligned receiver).

Summary of instructions

Here, you can see the instructions summarized once again. Take the time to understand them thoroughly once again. You will then be asked comprehension questions about the instructions. You can only continue with the experiment if you answer **all the comprehension questions correctly**.

- You are a bandit interacting with a gold miner and a partner.
- In each round, the gold miner either finds gold or does not find gold. Both states occur with a probability 50 percent.
- If the gold miner finds gold, he can tell or not tell the partner and bandit.
- If the gold miner does not find gold, the "no gold" message is automatically sent to the bandit and the partner.

- The gold miner cannot send different messages to the partner and the bandit. He sends the same message to both.
- Both the partner and the bandit earn more on average the more correct their guess is.
- The less the bandit believes that the gold miner has found gold, and the closer the partner's guess is to the truth, the more the gold miner earns on average.
- In each round, either the partner's guess or the bandit's guess is more important to the gold miner. Both conditions occur with a probability 50 percent.

☐ Details on payments

——— *new page* ———

Comprehension questions

Please answer the following 5 questions. If you answer [more than] one question incorrectly, you will be excluded from the experiment.

——— *new page* ———

Question 1

What's your role throughout the entire experiment?

Reply options:

- *Gold miner*
- *Bandit*
- *Partner*

——— *new page* ———

Question 2

The gold miner can send the message "no gold" also in case he did find gold.

Reply options:

- *True*
- *False*

——— *new page* ———

Question 3

Both the partner and the bandit earn more money on average the more accurately they guess whether the gold miner has found gold or not.

Reply options:

- *True*

- *False*

new page

Question 4

The gold miner wants that the partner's guess partner is

Reply options:

- *As high as possible*
- *As low as possible*
- *As correct as possible*

new page

Question 5

The gold miner wants that the bandit's guess is

Reply options:

- *As high as possible*
- *As low as possible*
- *As correct as possible*

new page

You have successfully answered the comprehension questions.

Congratulations, you have successfully answered the comprehension questions. Click Next so that you can be assigned to a group and begin the actual experiment.

new page

Please wait

Please wait until the other participants have finished this part of the experiment. You will now be divided into groups.

new page

III Rounds

Example for partner (aligned receiver).

Payment of gold miner in this round

In this round, your guess is more important for the gold miner than the bandit's guess.

<graphic illustration of this round's alignment>

You can see the exact payment when you click on the button below.

☐ Display role and payments

sender screen

You found gold.

You found gold. Do you want to tell this to the partner and the bandit?

Reply options:

- *"no gold"*
- *"gold"*

☐ Display role and payments

or

You did not find gold.

You did not find gold. Thus, the message "no gold" is sent automatically to both the partner and the bandit.

☐ Display role and payments

receiver screen

You guess

If the gold miner finds gold, he can either send the message "gold" or "no gold".

If he doesn't find any gold, he always sends you the "no gold" message automatically.

In case you receive the **"no gold" message**, how likely do you think it is that the gold miner found gold?

<slider 0,1...,100(%)>

You may be unaccustomed to expressing your beliefs on whether the gold miner has found gold as a percentage. So here are some examples of what these percentages roughly mean.

- 2 or 5 percent means something like: "It's very unlikely that the gold miner found gold."
- 18 percent means something like: "It's relatively unlikely that the gold miner found gold."
- 47 or 52 percent means something like: "It's about equally likely that the gold miner found gold and that he found no gold."
- 83 percent means something like: "It's relatively likely that the gold miner found gold."
- 95 or 98 percent means something like: "It's very likely that the gold miner found gold."

The gold miner can only send you the message "gold " if he has actually found gold. Therefore, you can then be sure that he has actually found gold. For this reason, your guess for this case is already logged in and you cannot adjust the slider.

<slider set to 100(%)>

☐ Display role and payments

new page

Please wait

Please wait a moment for the other participants in your group.

sender screen

What are the guesses of the partner and of the bandit?

You already know whether you have found gold or not. If you have found gold, you have decided between the messages “gold” and “no gold”. However, the partner and the bandit do not have all this knowledge yet.

The partner and bandit have each indicated how likely they think it is that you found gold when they receive the message “no gold”. Now, your task is to guess these guesses.

You can win another prize of 2 euros at this point. One of your guesses will be randomly selected. If it is close enough (not more than 5 percentage points away) to the actual guess of the partner or bandit, you win the prize.

In case you send the message “no gold” how likely does the partner think it is that you found gold?

<slider 0,1...,100(%)>

In case you send the message “no gold” how likely does the bandit think it is that you found gold?

<slider 0,1...,100(%)>

☐ Display role and payments

receiver screen

The gold miner’s message and your guess

The gold miner sent you **the message “no gold”**. In this case, you believe with probability <previously stated belief> that the gold miner found gold.

sender screen

Your payment of this round (1/8)

Round	(1/8) is now done.
Your role throughout the entire experiment	Gold miner
Did you find gold?	No gold. You did not find any gold.
Your message to the partner and the bandit	“no gold” You couldn’t say anything other than “no gold“ because you didn’t find any gold.
The partner’s guess in this case (guess_partner)	0%
The bandit’s guess in this case (guess_bandit)	50%
Whose guess was more important to you in this round?	The partner’s guess
Amount of lottery tickets (max. 400) for the prize of 8 euros you earned	350
Prize of 2 euros	You won it

After you click Next, the next round begins.

It is again randomly determined whether you will find gold or not.

It is also randomly determined whether the bandit’s guess or the partner’s guess is more important to you.

You start again with an urn of 400 rivets for the prize of 8 euros and can again win the additional prize of 2 euros.

☐ Display role and payments

receiver screen

Example for bandit (non-aligned receiver).

Your payment of this round (1/8)

Round	(1/8) is now done.
Your role throughout the entire experiment	Bandit
Gold miner's message	"no gold"
Your guess in this case	50%
Did the gold miner find gold?	The gold miner found gold. The gold miner told you that he did not find gold.
Whose guess was more important to the gold miner in this round?	Your guess
Amount of lottery tickets (max. 400) for the prize of 8 euros you earned	300

After you click Next, the next round begins.

It is again randomly determined whether the gold miner finds gold or not.

It is also randomly determined whether the bandit's guess or the partner's guess is more important to him.

You start again with an urn with 400 rivets.

☐ Display role and payments

new page

Please wait

Please wait until the other participants have also finished this round. You will now be assigned to new groups.

new page

7 more rounds.

IV Questionnaire

What was the role you were in throughout the entire experiment?

The last round of the experiment has been completed. What was your role in the entire experiment?

Reply options:

- *Gold miner*
- *Bandit*
- *Partner*

receiver screen

If applicable.

You have indicated one or more times that you believe with a probability of more than 50% that gold has been found when the gold miner sends you the message "no gold found". Why did you indicate this?

sender screen

If applicable.

You have indicated one or more times that you expect the partner and the bandit to react differently to the "no gold" message. You have indicated that the partner and the bandit think it is differently likely that you have found gold after they receive the same message from you. Why did you indicate this? Why do you think the partner and the bandit would have different guesses?

new page

Personal details (1/2)

The actual experiment is now finished. For filling out the following short questionnaire you will receive an additional 2 euros.

In politics, people often talk about "left" and "right." If you use a scale from 1 to 11, where would you classify yourself?

Reply options: *1=left, 2,...,6=don't know,7,...,11=right*

On average, how much time do you spend on social media (Twitter, Facebook, Instagram, etc.) each week?

Reply options:

- *I do not use social media.*
- *0-3 hours per week*
- *3-6 hours per week*
- *6-9 hours per week*
- *more than 9 hours per week*

Do you identify as part of a minority?

Reply options:

- *Yes*
- *No*
- *Rather not say*

new page

Personal details (2/2)

How old are you?

What's your gender?

What is your major?

What was your high school GPA?

What was your final grade in maths at high school?

How often have you participated in experiments?

Reply options:

- *Never*
- *1-2 times*
- *3-5 times*
- *More than 5 times*

How much income (including all financial support) do you have available in total per month?

Reply options:

- *Less than 1000 euros*
- *1000-1500 euros*
- *1500-2500 euros*
- *2500-3500 euros*
- *3500-5000 euros*
- *More than 5000 euros*
- *Rather not say*

——— *new page* ———

Feedback

If there is anything else you would like to tell us, please use this field:

——— *new page* ———

End of experiment

Thank you very much!

The entire experiment is now finished. You will receive a payment in the amount of 16 euros.

Please click on the following link to enter your payment details for your payment.

Appendix to Chapter 3

C.1 Tables

Table C.1: Differences in Intervention Over Time

Trial half	Number of villages	Control villages	Treatment villages
1	11	Appointments + letters scattered	Appointments + letters bunched
2	8	Appointments scattered + letters bunched	Appointments + letters bunched

Notes: The initially planned design was implemented for the first half of the trial, 11 clusters and $n = 7,365$. A weaker version where invitation letters are received simultaneously also in the control group was implemented for the second half of the trial, 8 clusters and $n = 13,354$.

Table C.2: Balance - Intervention

Variable	N	(1) Control Mean/SE	N	(2) Treatment Mean/SE	T-test Difference (1)-(2)
Went last time	10	0.564	9	0.542	0.022
		(0.015)		(0.008)	
Village population	10	8703.500	9	6829.556	1873.944
		(3028.442)		(1431.012)	
Number of invited women in village	10	1254.800	9	908.889	345.911
		(432.040)		(213.979)	
Age	10	59.376	9	59.519	-0.143
		(0.056)		(0.089)	
Distance to MMU (km)	10	6.704	9	8.653	-1.948
		(1.312)		(1.540)	
Covid 7d inc.	10	3.444	9	3.643	-0.199
		(0.306)		(0.829)	
Precipitation, in cm	10	0.418	9	0.108	0.310***
		(0.032)		(0.029)	
Avg. temperature (C)	10	13.810	9	10.556	3.254
		(2.334)		(3.140)	
Lead time	10	42.883	9	58.449	-15.566
		(5.932)		(13.826)	
School summer break	10	0.266	9	0.134	0.132
		(0.116)		(0.110)	

Notes: The value displayed for t-tests are the differences in the means across the groups. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table C.3: Results for First Invites - Dependent variable: Participation

	Simultaneous appointments		Simultaneous letters		Sim. appointments + letters	
	(1)	(2)	(3)	(4)	(5)	(6)
Treated	-0.066**	-0.077***	-0.042	-0.125***	-0.088**	-0.130***
	(0.028)	(0.024)	(0.027)	(0.024)	(0.036)	(0.038)
Lead time			-0.001	-0.001	-0.000	0.000
			(0.001)	(0.001)	(0.001)	(0.001)
Age	-0.018***	-0.016***	-0.017**	-0.011	-0.022***	-0.021***
	(0.004)	(0.004)	(0.006)	(0.007)	(0.004)	(0.004)
Precipitation, in cm		0.003		-0.012		-0.000
		(0.019)		(0.020)		(0.018)
Avg. temperature (C)		0.004*		-0.002		0.006
		(0.002)		(0.003)		(0.004)
Distance to MMU (km)		0.004		-0.003		0.002
		(0.003)		(0.003)		(0.002)
School break		-0.089*		-0.101**		-0.170***
		(0.046)		(0.032)		(0.037)
Covid 7d inc.		-0.012**		-0.011**		-0.021***
		(0.005)		(0.005)		(0.005)
Constant	1.372***	1.261***	1.400***	1.272***	1.598***	1.603***
	(0.222)	(0.244)	(0.348)	(0.374)	(0.200)	(0.295)
Observations	2274	2230	1434	1396	1341	1323
R^2	0.022	0.028	0.017	0.021	0.031	0.041
Clusters	19	19	10	10	15	15

Notes: Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors are clustered at the village level. Columns (1) – (3) compare all control villages (*Treated*= 0) with all treatment villages (*Treated*= 1). Columns (4) - (6) compare control villages of the first half of the trial (*Treated*= 0) to control villages of the second half of the trial (*Treated*= 1). Columns (7) - (9) compare control villages of the first half of the trial (*Treated*= 0) with all treatment villages (*Treated*= 1).

Table C.4: Survey 1

	(1) Perceived participation (binary)	(2) Perceived participation	(3) Car pooling	(4) Education	(5) Migration status
Treatment	0.038	0.059	0.000	0.010	-0.070**
	(0.027)	(0.091)	(0.010)	(0.046)	(0.024)
Constant	0.626***	3.281***	0.033***	2.992***	2.016***
	(0.015)	(0.036)	(0.009)	(0.039)	(0.016)
Observations	849	550	908	848	895
R^2	0.002	0.001	0.000	0.000	0.002
Clusters	11	11	11	11	11

Notes: Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors are clustered at the village level. *Perceived participation (binary)* takes on the value of 0 if and only if the respondent states that she has no idea how many of her acquaintances are participating in the program and 1 otherwise. *Perceived participation* is defined only for women for who *perceived participation (binary)* = 1. It takes on values from 0 to 4, where 0 means that a respondent believes almost no one of her acquaintances to participate and with step sizes of 1/4 of acquaintances, the maximum value of 4 means that she believes almost all of her acquaintances to participate. *Car pooling* is a binary variable taking on the value of 1 if a respondent was carpooling in order to reach the check-up site. *Education* is constructed from an 8-point Likert scale with higher values corresponding to higher levels of education. The variable *migration status* takes on the value of 1 if the respondent was born in the village she currently resides in and a maximum value of 4 if she was born outside of Germany.

Table C.5: Survey 2

	(1) Car pooling	(2) Education	(3) Migration status
Treatment	0.008	-0.494	-0.171
	(0.006)	(0.367)	(0.131)
Constant	0.000**	3.862***	2.074***
	(0.000)	(0.358)	(0.106)
Observations	627	627	627
R^2	0.005	0.011	0.010
Clusters	8	8	8

Notes: Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors are clustered at the village level. *Car pooling* is a binary variable taking on the value of 1 if a respondent was carpooling in order to reach the check-up site. *Education* is constructed from an 8-point Likert scale with higher values corresponding to higher levels of education. The variable *migration status* takes on the value of 1 if the respondent was born in the village she currently resides in and a maximum value of 4 if she was born outside of Germany.

Table C.6: Survey 2 - Conversations

	(1) Belief in talking openly	(2) Pre belief in talking openly	(3) Post belief in talking openly
Treatment	0.028^{**}	0.023	0.048
	(0.008)	(0.014)	(0.067)
Constant	0.867^{***}	0.871^{***}	0.848^{***}
	(0.006)	(0.004)	(0.024)
Observations	627	519	108
R^2	0.002	0.001	0.004
Clusters	8	8	8

Notes: Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors are clustered at the village level. *Belief in talking openly* is a binary variable taking on the value of 1 if and only if the respondent states that she believes women are talking openly about their participation in the program and 0 otherwise. Column (2) restricts the analysis of *belief in talking openly* to responses that were collected prior to the respondent's appointment. Column (3) restricts the analysis of *belief in talking openly* to responses that were collected after the respondent's appointment.

Table C.7: Survey 2 - Perceptions

	(1) Perceived participation (binary)	(2) Pre perceived participation (binary)	(3) Post perceived participation (binary)	(4) Perceived participation
Treatment	0.076	0.056	0.180^{*}	-0.169
	(0.048)	(0.053)	(0.080)	(0.092)
Constant	0.595^{***}	0.608^{***}	0.544^{***}	3.289^{***}
	(0.044)	(0.049)	(0.061)	(0.040)
Observations	627	519	108	391
R^2	0.006	0.003	0.026	0.007
Clusters	8	8	8	8

Notes: Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors are clustered at the village level. *Perceived participation (binary)* takes on the value of 0 if and only if the respondent states that she has no idea how many of her acquaintances are participating in the program and 1 otherwise. Column (2) restricts the analysis of *perceived participation (binary)* to responses that were collected prior to the respondent's appointment. Column (3) restricts the analysis of *perceived participation (binary)* to responses that were collected after the respondent's appointment. *Perceived participation* is defined only for women for who *perceived participation (binary)* $= 1$. It takes on values from 0 to 4, where 0 means that a respondent believes almost no one of her acquaintances to participate, and with step sizes of $1/4$ of acquaintances, the maximum value of 4 means that she believes almost all of her acquaintances to participate.

Table C.8: Peer Shares: Descriptive Statistics

Relative timing criterion	Peers	N	Mean	SD	p25	p75
Same day	d <50m	10110	0.054	0.137	0.000	0.000
	d <200m	10750	0.051	0.049	0.020	0.071
	d <500m	10793	0.050	0.032	0.032	0.063
	d <1000m	10813	0.050	0.029	0.035	0.059
Within 2 days	d <50m	10110	0.159	0.219	0.000	0.250
	d <200m	10750	0.158	0.089	0.100	0.203
	d <500m	10793	0.158	0.067	0.116	0.191
	d <1000m	10813	0.157	0.062	0.120	0.188
Up to today	d <50m	10110	0.513	0.367	0.000	0.800
	d <200m	10750	0.514	0.290	0.205	0.713
	d <500m	10793	0.512	0.284	0.209	0.705
	d <1000m	10813	0.512	0.282	0.208	0.702
The week before	d <50m	10110	0.205	0.243	0.000	0.250
	d <200m	10750	0.202	0.113	0.069	0.218
	d <500m	10793	0.201	0.093	0.083	0.210
	d <1000m	10813	0.200	0.088	0.085	0.208
7+ days earlier (Placebo)	d <50m	10110	0.308	0.347	0.000	0.545
	d <200m	10750	0.311	0.281	0.000	0.550
	d <500m	10793	0.311	0.275	0.023	0.564
	d <1000m	10813	0.312	0.273	0.029	0.566
7+ days later (Placebo)	d <50m	10110	0.329	0.339	0.000	0.571
	d <200m	10750	0.331	0.264	0.077	0.538
	d <500m	10793	0.332	0.257	0.090	0.538
	d <1000m	10813	0.332	0.256	0.095	0.534
Same day	k <0.5 years	11043	0.113	0.105	0.041	0.153
	k <1 year	11043	0.101	0.082	0.043	0.137
	k <2 years	11043	0.086	0.060	0.046	0.109
Within 2 days	k <0.5 years	11043	0.289	0.204	0.135	0.400
	k <1 year	11043	0.278	0.185	0.137	0.383
	k <2 years	11043	0.254	0.146	0.145	0.349
Up to today	k <0.5 years	11043	0.564	0.283	0.200	0.695
	k <1 year	11043	0.560	0.278	0.222	0.703
	k <2 years	11043	0.554	0.271	0.238	0.699
The week before	k <0.5 years	11043	0.321	0.203	0.067	0.328
	k <1 year	11043	0.312	0.189	0.077	0.325
	k <2 years	11043	0.297	0.168	0.087	0.315
7+ days earlier (Placebo)	k <0.5 years	11043	0.243	0.278	0.000	0.357
	k <1 year	11043	0.248	0.277	0.000	0.361
	k <2 years	11043	0.257	0.273	0.019	0.364
7+ days later (Placebo)	k <0.5 years	11043	0.229	0.226	0.049	0.337
	k <1 year	11043	0.231	0.223	0.055	0.334
	k <2 years	11043	0.236	0.222	0.060	0.344

Notes: This table presents descriptive characteristics of the *peer shares* values from section 3.5. The shares are constructed according to the set of peers and the share of them that fulfill a relative timing criterion (for an illustrative example, see Figure 3.2). The number of women, the mean of the peer share, the standard deviation, the 25th percentile, and the 75th percentile are reported. Deviations in N result from women with 0 pears according to the peer criterion.

Table C.9: Peer Shares Estimation - Criterion: Same Day

	(1) 50m	(2) 200m	(3) 500m	(4) 1000m
Share	-0.029	-0.046	-0.011	0.065
	(0.027)	(0.073)	(0.12)	(0.13)
Went last time	0.65***	0.66***	0.65***	0.66***
	(0.0074)	(0.0071)	(0.0071)	(0.0071)
n within distance	-0.0041***	-0.00040**	-0.00013***	-0.000025
	(0.0011)	(0.00017)	(0.000045)	(0.000018)
Constant	0.27***	0.26***	0.27***	0.25***
	(0.048)	(0.048)	(0.048)	(0.048)
Village FE + Controls	Yes	Yes	Yes	Yes
Observations	10110	10750	10793	10813

Notes: Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. This table and the subsequent tables C.10 to C.20 present results for estimations of Equation 3.3. The peer criterion is given in the respective column header and the relative timing criterion is presented in the table title.

Table C.10: Peer Shares Estimation - Criterion: Within 2 Days

	(1) 50m	(2) 200m	(3) 500m	(4) 1000m
Share	-0.0018	-0.021	-0.036	-0.061
	(0.017)	(0.042)	(0.057)	(0.063)
Went last time	0.65***	0.66***	0.65***	0.66***
	(0.0074)	(0.0071)	(0.0071)	(0.0071)
n within distance	-0.0041***	-0.00040**	-0.00013***	-0.000024
	(0.0011)	(0.00017)	(0.000045)	(0.000018)
Constant	0.27***	0.26***	0.28***	0.26***
	(0.048)	(0.048)	(0.048)	(0.048)
Village FE + Controls	Yes	Yes	Yes	Yes
Observations	10110	10750	10793	10813

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: See Table C.9 Notes.

Table C.11: Peer Shares Estimation - Criterion: Up to Today

	(1) 50m	(2) 200m	(3) 500m	(4) 1000m
Share	0.0059	0.0025	0.0058	0.0031
	(0.012)	(0.016)	(0.016)	(0.017)
Went last time	0.65***	0.66***	0.65***	0.66***
	(0.0074)	(0.0072)	(0.0072)	(0.0072)
n within distance	-0.0041***	-0.00040**	-0.00013***	-0.000025
	(0.0011)	(0.00017)	(0.000045)	(0.000018)
Constant	0.26***	0.26***	0.26***	0.24***
	(0.055)	(0.061)	(0.062)	(0.062)
Village FE + Controls	Yes	Yes	Yes	Yes
Observations	10110	10750	10793	10813

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: See Table C.9 Notes.

Table C.12: Peer Shares Estimation - Criterion: 1 to 7 Days Before

	(1) 50m	(2) 200m	(3) 500m	(4) 1000m
Share	0.050***	0.14***	0.18***	0.19***
	(0.018)	(0.035)	(0.042)	(0.045)
Went last time	0.65***	0.66***	0.66***	0.66***
	(0.0073)	(0.0071)	(0.0071)	(0.0071)
n within distance	-0.0042***	-0.00040**	-0.00013***	-0.000025
	(0.0011)	(0.00017)	(0.000045)	(0.000018)
Constant	0.26***	0.23***	0.23***	0.21***
	(0.049)	(0.049)	(0.049)	(0.049)
Village FE + Controls	Yes	Yes	Yes	Yes
Observations	10110	10750	10793	10813

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: See Table C.9 Notes.

Table C.13: Peer Shares Estimation - Criterion: 7+ Days Earlier (Placebo)

	(1) 50m	(2) 200m	(3) 500m	(4) 1000m
Share	-0.017	-0.025	-0.020	-0.022
	(0.013)	(0.017)	(0.017)	(0.018)
Went last time	0.65***	0.66***	0.66***	0.66***
	(0.0074)	(0.0072)	(0.0072)	(0.0072)
n within distance	-0.0041***	-0.00040**	-0.00013***	-0.000024
	(0.0011)	(0.00017)	(0.000045)	(0.000018)
Constant	0.30***	0.31***	0.31***	0.30***
	(0.055)	(0.060)	(0.061)	(0.061)
Village FE + Controls	Yes	Yes	Yes	Yes
Observations	10110	10750	10793	10813

Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
Notes: See Table C.9 Notes.

Table C.14: Peer Shares Estimation - Criterion: 7+ Days Later (Placebo)

	(1) 50m	(2) 200m	(3) 500m	(4) 1000m
Share	-0.0070	-0.00091	0.0013	0.0012
	(0.012)	(0.017)	(0.018)	(0.018)
Went last time	0.65***	0.66***	0.66***	0.66***
	(0.0074)	(0.0072)	(0.0072)	(0.0072)
n within distance	-0.0041***	-0.00040**	-0.00013***	-0.000025
	(0.0011)	(0.00017)	(0.000045)	(0.000018)
Constant	0.26***	0.26***	0.27***	0.25***
	(0.050)	(0.052)	(0.053)	(0.053)
Village FE + Controls	Yes	Yes	Yes	Yes
Observations	10110	10750	10793	10813

Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
Notes: See Table C.9 Notes.

Table C.15: Peer Shares Estimation - Criterion: Same Day

	(1) 0.5 years	(2) 1 year	(3) 2 years
Share	0.043	0.018	0.056
	(0.044)	(0.057)	(0.080)
Went last time	0.66***	0.66***	0.66***
	(0.0074)	(0.0074)	(0.0073)
n within age diff.	0.000062	-0.0000030	0.000021
	(0.00019)	(0.00010)	(0.000048)
Constant	0.18*	0.22**	0.18*
	(0.10)	(0.11)	(0.10)
Village FE + Controls	Yes	Yes	Yes
Observations	10817	10817	10817

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: See Table C.9 Notes.

Table C.16: Peer Shares Estimation - Criterion: Within 2 Days

	(1) 0.5 years	(2) 1 year	(3) 2 years
Share	-0.0059	-0.018	-0.017
	(0.021)	(0.024)	(0.029)
Went last time	0.66***	0.65***	0.65***
	(0.0073)	(0.0074)	(0.0074)
n within age diff.	0.00000086	-0.000032	0.0000059
	(0.00019)	(0.00010)	(0.000049)
Constant	0.24**	0.28***	0.23**
	(0.100)	(0.11)	(0.10)
Village FE + Controls	Yes	Yes	Yes
Observations	10817	10817	10817

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: See Table C.9 Notes.

Table C.17: Peer Shares Estimation - Criterion: Up to Today

	(1) 0.5 years	(2) 1 year	(3) 2 years
Share	-0.0071	-0.00012	0.0036
	(0.013)	(0.014)	(0.014)
Went last time	0.66***	0.66***	0.66***
	(0.0072)	(0.0072)	(0.0072)
n within age diff.	0.000015	-0.000011	0.000013
	(0.00018)	(0.000097)	(0.000048)
Constant	0.22**	0.24**	0.20**
	(0.089)	(0.094)	(0.095)
Village FE + Controls	Yes	Yes	Yes
Observations	10817	10817	10817

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: See Table C.9 Notes.

Table C.18: Peer Shares Estimation - Criterion: 1 to 7 Days Before

	(1) 0.5 years	(2) 1 year	(3) 2 years
Share	0.048**	0.062**	0.087***
	(0.023)	(0.024)	(0.026)
Went last time	0.66***	0.66***	0.66***
	(0.0071)	(0.0071)	(0.0071)
n within age diff.	0.000036	0.000011	0.000035
	(0.00018)	(0.000097)	(0.000048)
Constant	0.21**	0.21**	0.15
	(0.090)	(0.095)	(0.096)
Village FE + Controls	Yes	Yes	Yes
Observations	10817	10817	10817

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: See Table C.9 Notes.

Table C.19: Peer Shares Estimation - Criterion: 7+ Days Earlier (Placebo)

	(1) 0.5 years	(2) 1 year	(3) 2 years
Share	-0.027*	-0.021	-0.021
	(0.015)	(0.015)	(0.015)
Went last time	0.66***	0.66***	0.66***
	(0.0073)	(0.0073)	(0.0073)
n within age diff.	0.000034	-0.0000031	0.000015
	(0.00018)	(0.000097)	(0.000048)
Constant	0.22**	0.24**	0.21**
	(0.089)	(0.094)	(0.095)
Village FE + Controls	Yes	Yes	Yes
Observations	10817	10817	10817

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: See Table C.9 Notes.

Table C.20: Peer Shares Estimation - Criterion: 7+ Days Later (Placebo)

	(1) 0.5 years	(2) 1 year	(3) 2 years
Share	0.015	0.013	0.0075
	(0.017)	(0.017)	(0.017)
Went last time	0.66***	0.66***	0.66***
	(0.0071)	(0.0071)	(0.0071)
n within age diff.	-0.0000044	-0.000021	0.0000098
	(0.00019)	(0.000098)	(0.000048)
Constant	0.23***	0.25***	0.21**
	(0.090)	(0.095)	(0.095)
Village FE + Controls	Yes	Yes	Yes
Observations	10817	10817	10817

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: See Table C.9 Notes.

C.2 Figures

Central Office Xxxxxxx | Any Street XX | XXXXX Any City

Ms.
<first name> <last name>
<street> <house numer>
<postcode> <city>

<date>

EARLY DETECTION OF BREAST CANCER:
OFFER OF AN EXAMINATION AS PART OF THE MAMMOGRAPHY SCREENING PROGRAMME

Dear Ms <last name>

in Germany, women between the ages of 50 and 69 have the opportunity to participate in the Mammography Screening Programme for the Early Detection of Breast Cancer every two years. The goal is to be able to better treat breast cancer through early discovery and to reduce mortality from breast cancer.

As the "Central Office", we have the mission to inform you about this and to invite you to the mammography examinations. Gladly, we suggest the following appointment for a mammography examination:

<date> at <time>
<mammography-facility>
<street> in < postcode > <city>
<Placeholder for directions to the Mammobil>

If you would like a different appointment, have questions or would like to cancel, you can contact us at <phone> or by e-mail to <email>, by fax to <fax> or letter.

Important is: Participation in the mammography screening is voluntary. Like all early detection screenings, the mammography has advantages and disadvantages. A brochure is enclosed with this invitation to support you in your personal decision for or against participation. You can also find further information on the Internet at https://www.g-ba.de/entscheidungshilfe-mammographie.de.

You have the right to a personal consultation with a doctor from the mammography programme. In this conversation you can have the advantages and disadvantages explained to you in detail and open questions can be answered. There are usually no doctors present during the mammography examination itself.

If you wish to have such an interview, you must make a separate appointment for this before the examination. Please contact us as the Central Office for this.

You can also take part in the early detection screening without personal consultation. In this case, please bring the enclosed signed declaration on the waiver of personal consultation.

Further information on participation or cancellation can be found on the back of this letter.

With kind regards

Please turn →

IF YOU DO NOT WISH TO PARTICIPATE

You are entitled to this offer every two years. If you do not wish to participate this time, we will contact you again in two years.

If you do not wish to receive any further invitations, please inform us in text form by fax to <fax>, by letter to <address> <postcode> <city> or by e-mail to <email>. If you change your decision later, please inform us. We will then send you a new invitation.

If you do not participate, you will not suffer any disadvantages in terms of health insurance and care. Even if you should develop breast cancer at some point, your health insurance will of course cover the treatment costs.

INFORMATION ON PARTICIPATION - PLEASE NOTE IN ADVANCE

The costs of the examination are covered by your statutory health insurance. A referral is not necessary. If you are privately insured, please clarify the cost absorption with your insurance company in advance.

- ➔ The mammography screening is for women who have no signs of breast disease.
- ➔ If you have already had a mammography screening examination within the last 22 months or have had a mammography within the last 12 months for other reasons (e.g. after breast cancer), please let us know in advance.
- ➔ If you need help or are dependent on a wheelchair, please contact us in advance, as the central office.

ON THE DAY OF THE EXAMINATION - PLEASE NOTE

Please bring your insurance card, this invitation letter and the completed questionnaire. If you do not want a personal consultation, also the signed waiver.

Please do not use powder, deodorant or cream on the chest and underarm area on the day of the examination as this may interfere with the X-ray images.

THE RESULT OF THE MAMMOGRAM

Mammography is used to look for abnormalities that indicate breast cancer. You will usually be informed by the mammography unit within seven working days whether such abnormalities have been found or not. If abnormalities are found, this does not mean that it has to be breast cancer. In most cases, the suspicion can be disproved. However, further examinations are necessary. You will then receive another invitation.

PRIVACY

Your address was provided to us by your municipality in accordance with the legal requirements for data protection. The protection of your data is ensured at all times. Your examination results are only available at the mammography facility and are subject to medical confidentiality. Further information on the use of your data can be found in the enclosed brochure.

DECLARATION ON THE WAIVER OF THE PERSONAL CONSULTATION

I have been informed of the main advantages and disadvantages of the mammography screening programme by the enclosed documents and waive my right to an additional personal consultation with a doctor of the programme before the examination.

<first name> <last name>, born on <date of birth>

Date | Signature

Figure C.1: Invitation Letter - English Translation

Notes: The original German version is available at `https://www.mammo-programm.de/de/info-deutsch`.

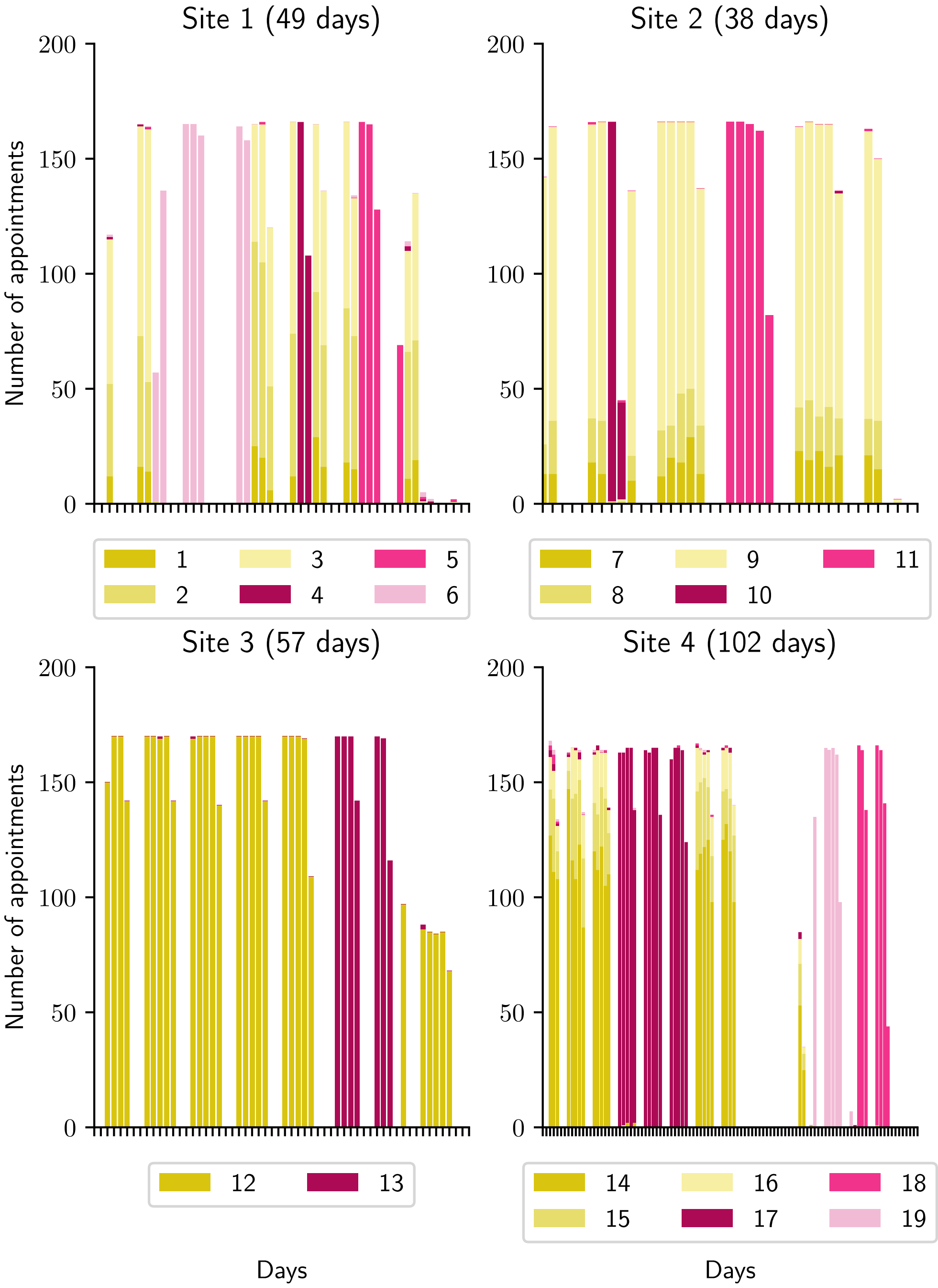


Figure C.2: Manipulation of Initial Appointments

Notes: Control villages are depicted in green shades, and treatment villages are depicted in red shades. The average standard deviation of initial appointment dates weighted by village size in the treatment (control) group is 5.24 (14.71).

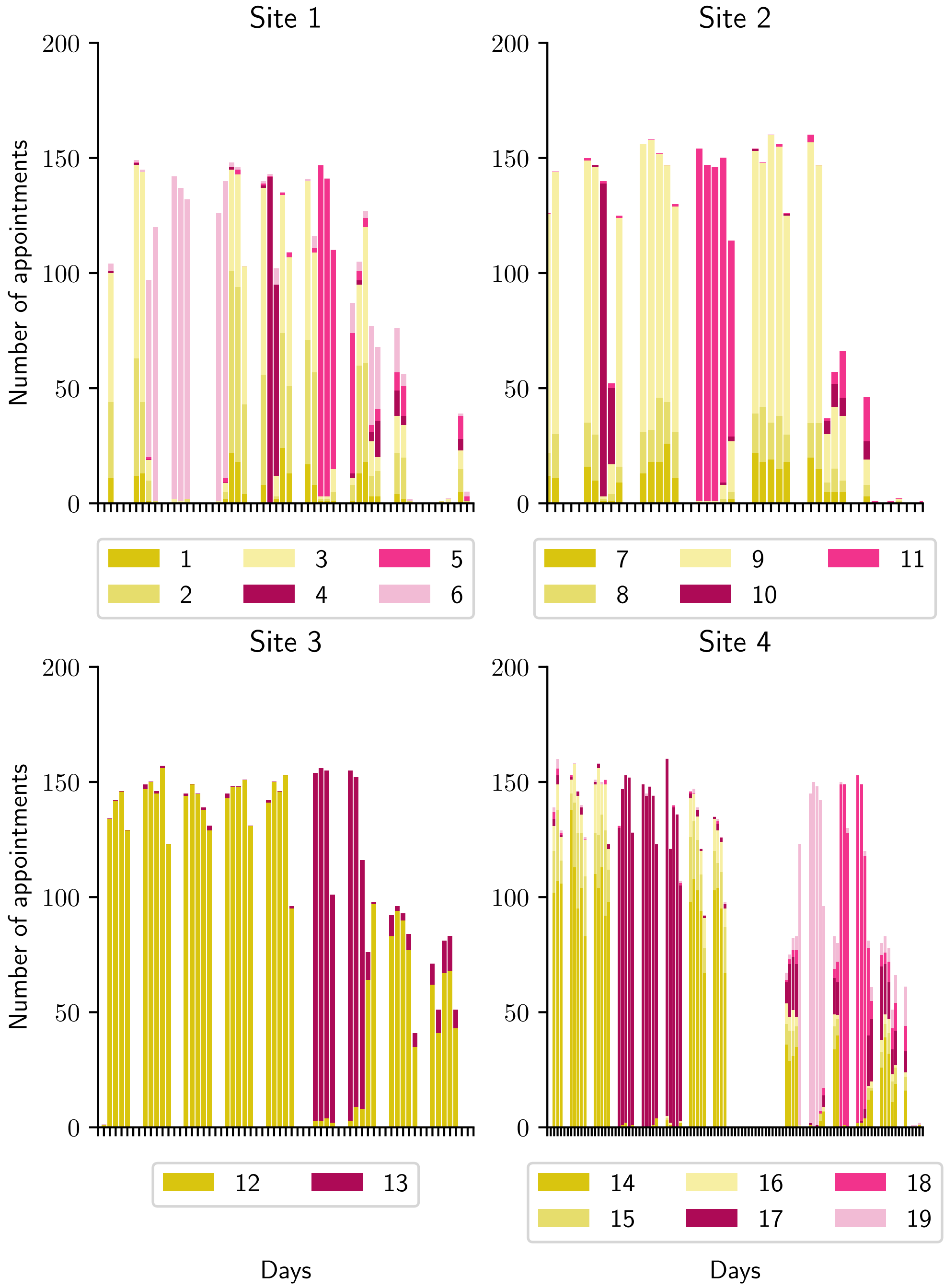


Figure C.3: Manipulation of Final Appointments

Notes: Control villages are depicted in green shades, and treatment villages are depicted in red shades. The average standard deviation of final appointment dates weighted by village size in the treatment (control) group is 15.13 (21.01).

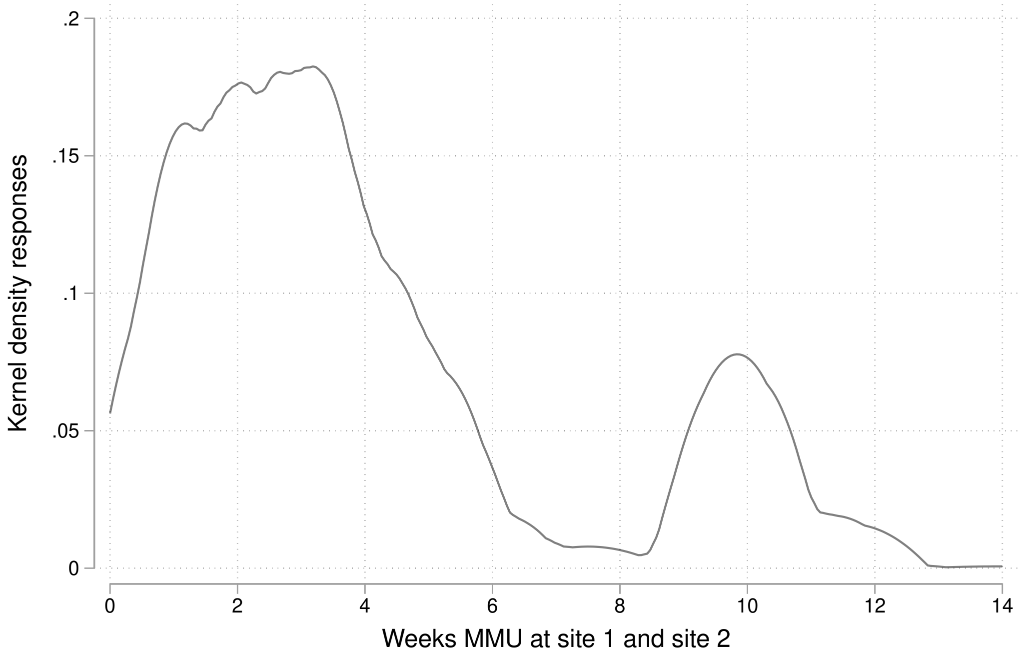


Figure C.4: Survey 1 Responses Over Time

Notes: The 946 survey 1 responses were all collected at sites 1 and 2. Most of the answers were collected in the first five weeks.

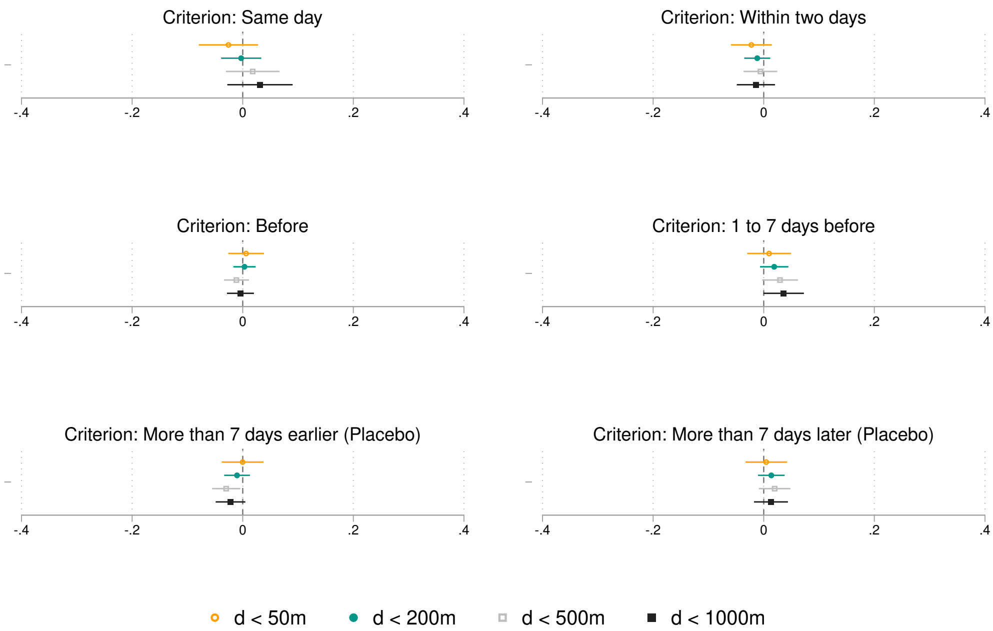


Figure C.5: Peer Shares by Geographic Distance, Women with <0.5 Years Age Difference

Notes: This figure presents resulting β_1 from estimations of Equation 3.3 which can be interpreted as the change in participation in % given an increase in $share_i$ by 1%. Each point estimate and its 95% confidence bands represent one regression that includes village-fixed effects and controls for previous participation, the size of the peer group, distance to the mammography unit, and age. The header of each panel indicates the relative timing criterion (see Table 3.5). Each bar per panel represents a different peer criterion, here by distance in meters (see Table 3.4) and age difference, as only women with <0.5 years age difference are considered.

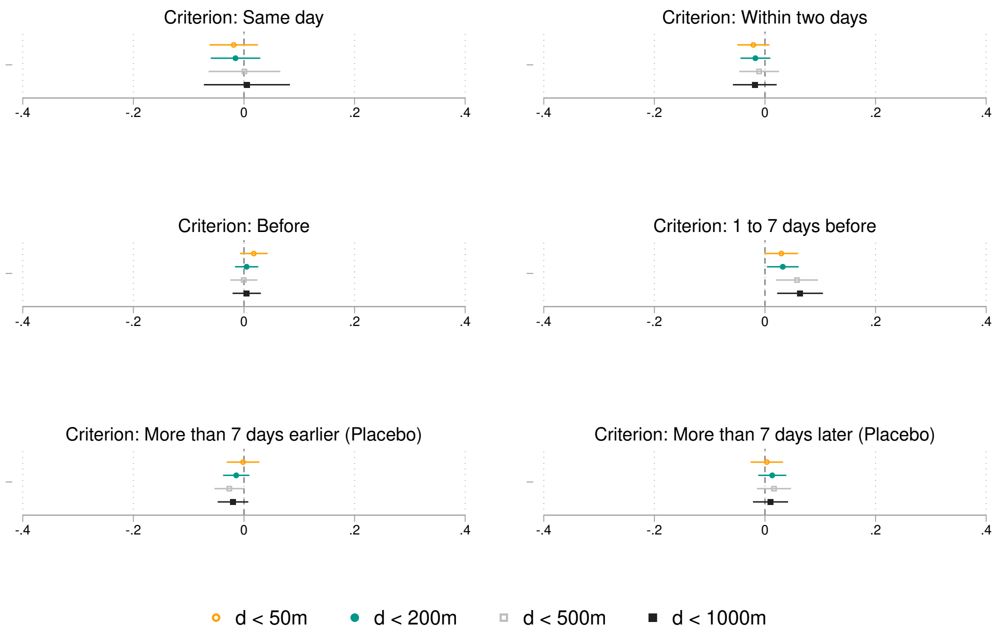


Figure C.6: Peer Shares by Geographic Distance, Women with <1 Year Age Difference

Notes: This figure presents resulting β_1 from estimations of Equation 3.3 which can be interpreted as the change in participation in % given an increase in $share_i$ by 1%. Each point estimate and its 95% confidence bands represent one regression that includes village-fixed effects and controls for previous participation, the size of the peer group, distance to the mammography unit, and age. The header of each panel indicates the relative timing criterion (see Table 3.5). Each bar per panel represents a different peer criterion, here by distance in meters (see Table 3.4) and age difference, as only women with <1 year age difference are considered.

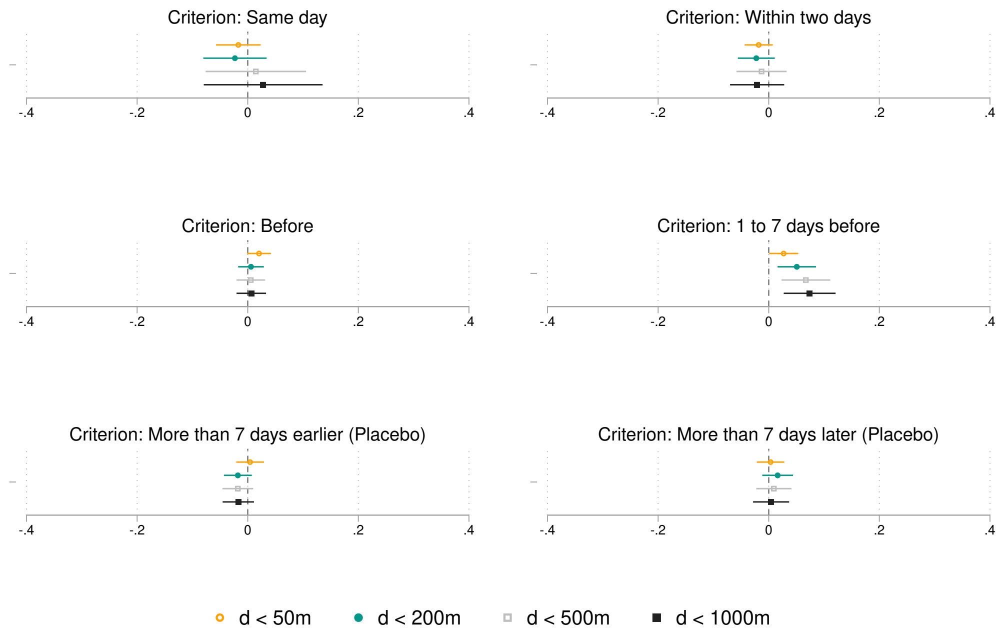


Figure C.7: Peer Shares by Geographic Distance, Women with <2 Year Age Difference

Notes: This figure presents resulting β_1 from estimations of Equation 3.3 which can be interpreted as the change in participation in % given an increase in $share_i$ by 1%. Each point estimate and its 95% confidence bands represent one regression that includes village-fixed effects and controls for previous participation, the size of the peer group, distance to the mammography unit, and age. The header of each panel indicates the relative timing criterion (see Table 3.5). Each bar per panel represents a different peer criterion, here by distance in meters (see Table 3.4) and age difference, as only women with <2 years of age difference are considered.

C.3 Survey Material

C.3.1 Paper based survey

I Preliminaries

We are scientists at the Ludwig-Maximilians-University Munich and do research on the topic of breast cancer screening. We would be very pleased if you answered the questions below. Please be assured that your data will be treated confidentially and will not be traced back to you. By consenting, you confirm that you are of legal age and that you know that filling out the questionnaire is voluntary.

II Residence

Question 1: Which village do you reside in?

Question 2: Where were you born?
Reply options:

- *In my current place of residence.*
- *In [federal state], but not in my current place of residence.*
- *In Germany but outside of [federal state].*
- *Outside of Germany.*

III Mammography

Question 1: Have you ever gone for breast cancer screening?
Reply options (multiple answers possible):

- *Yes, like today, as part of the mammography screening program.*
- *Yes, in a gynecology or radiology office.*
- *No.*

Question 2: Would you like to go for breast cancer screening *more often, as often, or less often* than you have in the past?
Reply options: *more often, as often, less often*

Question 3: Do you have a rough idea of how many women in your circle of acquaintances participate in the mammography screening program?
Reply options: Relative shares *1, 0.75, 0.5, 0.25, 0, don't know*

Question 4: How did you get to the examination today?
Reply options:

- *Alone by car.*
- *In a carpool.*
- *By public transport.*
- *On foot or by bike.*

IV Demographics

Question 1: In which year were you born?

Question 2: What is your highest high school or college degree?

Reply options:

- *Without a general school leaving certificate.*
- *Without a general school leaving certificate.*
- *Secondary school diploma (Realschulabschluss).*
- *High school diploma or equivalent (Abitur).*
- *Apprenticeship.*
- *Study without a degree.*
- *Preliminary diploma (Vordiplom).*
- *Diploma (Diplom).*
- *Ph.D.*

V Closing Questions

Prompt 1: Please enter today's date (day, month).

Prompt 2: Space for general comments.

C.3.2 Online survey

I Preliminaries

Welcome to this survey. We are scientists from the University of Munich and do research on the topic of breast cancer prevention. We kindly ask you to answer the following questions. This shouldn't take more than 15 minutes of your time.

Please be assured that your data will be treated confidentially and cannot be traced back to you. By agreeing, you confirm that you are of legal age and understand that participation in the survey is voluntary.

II Context

You have been invited to breast cancer screening as part of the mammography screening program. (The invitation included the link to this survey.)

Question: Are you planning to go for breast cancer screening this year as part of the mammography screening program? Which is most applicable to you?
Reply options:

- *I plan to attend.*
- *I plan not to attend.*
- *I'm still undecided.*
- *I have already participated.*
- *I have not already participated.*

III Social Questions

Question 1: Have you ever gone for breast cancer screening?
Reply options (multiple answers possible):

- *Yes, as part of the mammography screening program.*
- *Yes, at the family doctor or in a gynecology or radiology practice.*
- *No.*

Question 2: Did you go for breast cancer screening in 2020 as part of the mammography screening program?
Reply options: *yes, no, don't remember*

Question 3: Would you say that women talk openly about participating in the mammography screening program?
Reply options: *yes, rather yes, rather no, no*

Question 4: Do you have a rough idea of how many women in your circle of acquaintances participate in the mammography screening program?
Reply options: Relative shares *1, 0.75, 0.5, 0.25, 0, don't know*

Question 5: Would you like to go to breast cancer screening *more often, as often or less often* than you have done in the past?
Reply options: *more often, as often, less often*

IV Questions on Participation

A-E is chosen depending on the answer in II.

IV.A Attendance Planned

Question 1: Why do you want to participate in the mammography screening program this year?

Question 2: If you are participating in the mammography screening program this year, how do you plan to get to the screening?
Reply options:

- *Alone by car.*
- *In a carpool.*
- *By public transport.*
- *On foot or by bike.*
- *Don't know.*

IV.B Attendance Not Planned

Question Why don't you want to participate in the mammography screening program this year?

IV.C Undecided

Question 1: Why are you still undecided whether or not to participate in the mammography screening program this year?

Question 2: If you are participating in the mammography screening program this year, how do you plan to get to the screening?

Reply options: See IV.A

IV.D Participated

Question 1: Why did you participate in the mammography screening program this year?

Question 2: How did you get to your examination within the mammography screening program?
Reply options: See IV.A

IV.E Didn't Participate

Question: Why didn't you participate in the mammography screening program this year?

V Social Potency

We now ask you to fill out a short personality questionnaire. If you don't have time for this, you can skip this part.

For each statement, please indicate to what extent it applies to you.
Reply options: *true, false*

- **Statement 1:** I am quite effective at talking people into things.
- **Statement 2:** I am very good at influencing people.
- **Statement 3:** I do not like to be the center of attention on social occasions.
- **Statement 4:** I do not enjoy trying to convince people of something.
- **Statement 5:** In most social situations, I like to have someone else take the lead.
- **Statement 6:** In social situations, I usually allow others to dominate the conversation.
- **Statement 7:** When it is time to make decisions, others usually turn to me.

VI Demographics

Question 1: Which village do you reside in?

Question 2: Where were you born?
Reply options:

- *In my current place of residence.*
- *In [federal state], but not in my current place of residence.*
- *In Germany but outside of [federal state].*
- *Outside of Germany.*

Question 3: Is German your mother tongue?

Reply options: *yes, no*

Question 4: What is your family status?

Reply options: *married, partnership, widowed, divorced, single*

Question 5: Do you have children?

Reply options: *yes, no*

Question 6: What is your highest school or college degree?

Reply options:

- *Without a general school leaving certificate.*
- *Without a general school leaving certificate.*
- *Secondary school diploma (Realschulabschluss).*
- *High school diploma or equivalent (Abitur).*
- *Apprenticeship.*
- *Study without a degree.*
- *Preliminary diploma (Vordiplom).*
- *Diploma (Diplom).*
- *Ph.D.*

Question 7: What is your husband's or partner's highest educational qualification?

Reply options: see the previous question

Question 8: Are you currently employed?

Reply options: *full time, part-time, no*

Question 9: Are you involved in your community? For example, in a club or in church?

Reply options: *yes, no*

Question 10: Please enter your date of birth (DD/MM/YYYY).

Bibliography

Acemoglu, Daron, Victor Chernozhukov, and Muhamet Yildiz. 2016. "Fragility of asymptotic agreement under Bayesian learning." *Theoretical Economics*, 11(1): 187–225.

Adena, Maja, Ruben Enikolopov, Maria Petrova, Veronica Santarosa, and Ekaterina Zhuravskaya. 2015. "Radio and the Rise of the Nazis in Prewar Germany." *Quarterly Journal of Economics*, 130(4): 1885–1939.

Alan, Sule, Gozde Corekcioglu, and Matthias Sutter. 2023. "Improving workplace climate in large corporations: A clustered randomized intervention." *Quarterly Journal of Economics*, 138(1): 151–203.

Ali, S. Nageeb, and Roland Bénabou. 2020. "Image versus Information: Changing Societal Norms and Optimal Privacy." *American Economic Journal: Microeconomics*, 12(3): 116–164.

Allcott, Hunt. 2011. "Social norms and energy conservation." *Journal of Public Economics*, 95(9): 1082–1095.

Allcott, Hunt, and Todd Rogers. 2014. "The Short-Run and Long-Run Effects of Behavioral Interventions: Experimental Evidence from Energy Conservation." *American Economic Review*, 104(10): 3003–3037.

Alsan, Marcella, and Sarah Eichmeyer. 2023. "Experimental Evidence on the Effectiveness of Non-Experts for Improving Vaccine Demand." *American Economic Journal: Economic Policy.*

Alsan, Marcella, Owen Garrick, and Grant Graziani. 2019. "Does Diversity Matter for Health? Experimental Evidence from Oakland." *American Economic Review*, 109(12): 4071–4111.

Andre, Peter, Teodora Boneva, Felix Chopra, and Armin Falk. 2021. "Fighting climate change: The role of norms, preferences, and moral values." CEPR Discussion Paper No. DP16343.

Babcock, Linda, George Loewenstein, Samuel Issacharoff, and Colin Camerer. 1995. "Biased judgments of fairness in bargaining." *American Economic Review*, 85(5): 1337–1343.

Banerjee, Abhijit Vinayak, Esther Duflo, Rachel Glennerster, and Dhruva Kothari. 2010. "Improving immunisation coverage in rural India: clustered randomised controlled evaluation of immunisation campaigns with and without incentives." *BMJ*, 340: c2220.

Battaglini, Marco, and Uliana Makarov. 2014. "Cheap talk with multiple audiences: An experimental analysis." *Games and Economic Behavior*, 83: 147–164.

Baysan, Ceren. 2022. "Persistent Polarizing Effects of Persuasion: Experimental Evidence from Turkey." *American Economic Review*, 112(11): 3528–3546.

Beaman, Lori, Ariel BenYishay, Jeremy Magruder, and Ahmed Mushfiq Mobarak. 2021. "Can Network Theory-Based Targeting Increase Technology Adoption?" *American Economic Review*, 111(6): 1918–1943.

Benartzi, Shlomo, Ehud Peleg, and Richard H. Thaler. 2013. "Choice Architecture and Retirement Saving Plans." In *The Behavioral Foundations of Public Policy*. 245–263. Princeton University Press.

Bernheim, B. Douglas. 1994. "A Theory of Conformity." *Journal of Political Economy*, 102(5): 841–877.

Bertoni, Marco, Luca Corazzini, and Silvana Robone. 2020. "The Good Outcome of Bad News: A Field Experiment on Formatting Breast Cancer Screening Invitation Letters." *American Journal of Health Economics*, 6(3): 372–409.

Bicchieri, Cristina, and Azi Lev-On. 2007. "Computer-mediated communication and cooperation in social dilemmas: an experimental analysis." *Politics, Philosophy & Economics*, 6(2): 139–255.

Biesheuvel, Cornelis, Stefanie Weige, and Walter Heindel. 2011. "Mammography Screening: Evidence, History and Current Practice in Germany and Other European Countries." *Breast Care*, 6(2): 104–109.

Borsky, Amanda, Chunliu Zhan, Therese Miller, Quyen Ngo-Metzger, Arlene S. Bierman, and David Meyers. 2018. "Few Americans receive all high-priority, appropriate clinical preventive services." *Health Affairs*, 37(6): 925–928.

Bouckaert, Nicolas, Anne C. Gielen, and Tom Van Ourti. 2020. "It runs in the family – Influenza vaccination and spillover effects." *Journal of Health Economics*, 74: 102386.

Braghieri, Luca. 2021. "Political Correctness, Social Image, and Information Transmission." Working Paper.

Braghieri, Luca. 2023. "Biased Decoding and the Foundations of Communication." Working Paper.

Brunnermeier, Markus K., and Jonathan A. Parker. 2005. "Optimal Expectations." *American Economic Review*, 95(4): 1092–1118.

Bursztyn, Leonardo, Alessandra L. González, and David Yanagizawa-Drott. 2020. "Misperceived Social Norms: Women Working Outside the Home in Saudi Arabia." *American Economic Review*, 110(10): 2997–3029.

Bursztyn, Leonardo, and David Y. Yang. 2022. "Misperceptions About Others." *Annual Review of Economics*, 14(1): 425–452.

Bursztyn, Leonardo, and Robert Jensen. 2017. "Social Image and Economic Behavior in the Field: Identifying, Understanding, and Shaping Social Pressure." *Annual Review of Economics*, 9(1): 131–153.

Bénabou, Roland, and Jean Tirole. 2011. "Laws and norms." NBER Working Paper No. 17579.

Cameron, A. Colin, Jonah B. Gelbach, and Douglas L. Miller. 2008. "Bootstrap-Based Improvements for Inference with Clustered Errors." *The Review of Economics and Statistics*, 90(3): 414–427.

Campos-Mercade, Pol, and Erik Wengström. 2020. "Threshold Incentives and Academic Performance." Working Paper.

Campos-Mercade, Pol, Armando N. Meier, Florian H. Schneider, Stephan Meier, Devin Pope, and Erik Wengström. 2021. "Monetary incentives increase COVID-19 vaccinations." *Science*, 374(6569): 879–882.

Canay, Ivan A., Andres Santos, and Azeem M. Shaikh. 2021. "The Wild Bootstrap with a "Small" Number of "Large" Clusters." *Review of Economics and Statistics*, 103(2): 346–363.

Carrieri, Vincenzo, and Ansgar Wübker. 2016. "Quasi-Experimental Evidence on the Effects of Health Information on Preventive Behaviour in Europe." *Oxford Bulletin of Economics and Statistics*, 78(6): 765–791.

Castro, Silvia, and Clarissa Mang. 2022. "Breaking the Silence - Group Discussions, Social Pressure, and the Adoption of Health Technologies." Working Paper.

Chauvin, Kyle P. 2023. "Unacknowledged Heterogeneity in Communication." Working Paper.

Chen, Daniel L, Martin Schonger, and Chris Wickens. 2016. "oTree—An open-source platform for laboratory, online, and field experiments." *Journal of Behavioral and Experimental Finance*, 9: 88–97.

Chen, M. Keith. 2013. "The Effect of Language on Economic Behavior: Evidence from Savings Rates, Health Behaviors, and Retirement Assets." *American Economic Review*, 103(2): 690–731.

Conlon, John, Malavika Mani, Gautam Rao, Matthew Ridley, and Frank Schilbach. 2021. "Learning in the Household." National Bureau of Economic Research w28844, Cambridge, MA.

Cooper, Russell, Douglas V. DeJong, Robert Forsythe, and Thomas W. Ross. 1992. "Communication in Coordination Games." *The Quarterly Journal of Economics*, 107(2): 739–771.

Crawford, Vincent P., and Joel Sobel. 1982. "Strategic Information Transmission." *Econometrica*, 50(6): 1431–1451. Publisher: [Wiley, Econometric Society].

Crémer, Jacques, Luis Garicano, and Andrea Prat. 2007. "Language and the Theory of the Firm." *Quarterly Journal of Economics*, 122(1): 373–407.

Dafni, Urania, Zoi Tsourti, and Ioannis Alatsathianos. 2019. "Breast Cancer Statistics in the European Union: Incidence and Survival across European Countries." *Breast Care*, 14(6): 344–353.

Dai, Hengchen, Silvia Saccardo, Maria A. Han, Lily Roh, Naveen Raja, Sitaram Vangala, Hardikkumar Modi, Shital Pandya, Michael Sloyan, and Daniel M. Croymans. 2021. "Behavioural nudges increase COVID-19 vaccinations." *Nature*, 597(7876): 404–409.

Deutsches Mammographie-Screening-Programm. 2022. "Jahresbericht Evaluation 2020." Kooperationsgemeinschaft Mammographie, Berlin.

Dilme, Francesc. 2018. "Optimal Languages." *SSRN Electronic Journal.* https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3331077.

Djourelova, Milena. 2023. "Persuasion through Slanted Language: Evidence from the Media Coverage of Immigration." *American Economic Review*, 113(3): 800–835.

Drago, Francesco, Friederike Mengel, and Christian Traxler. 2020. "Compliance Behavior in Networks: Evidence from a Field Experiment." *American Economic Journal: Applied Economics*, 12(2): 96–133.

Einav, Liran, Amy Finkelstein, Tamar Oostrom, Abigail Ostriker, and Heidi Williams. 2020. "Screening and Selection: The Case of Mammograms." *American Economic Review*, 110(12): 3836–3870.

Enke, Benjamin. 2020. "What You See Is All There Is." *Quarterly Journal of Economics*, 135(3): 1363–1398.

Enke, Benjamin, and Florian Zimmermann. 2019. "Correlation neglect in belief formation." *Review of Economic Studies*, 86(1): 313–332.

Erdmann, Friederike, Claudia Spix, Alexander Katalinic, Monika Christ, Juliane Folkerts, Jutta Hansmann, Kristine Kranzhöfer, Beatrice Kunz, Katrin Manegold, Andrea Penzkofer, Kornelia Treml, Grit Vollmer, Susanne Weg-Remers, Benjamin Barnes, Nina Buttmann-Schweiger, Stefan Dahm, Julia Fiebig, Manuela Franke, Ina Gurung-Schönfeld, Jörg Haberland, Maren Imhoff, Klaus Kraywinkel, Anne Starker, Petra von Berenberg-Gossler, and Antje Wienecke. 2021. "Krebs in Deutschland für 2017/2018." Robert Koch-Institut 13, Berlin.

Esguerra, Emilio, Leonhard Vollmer, and Johannes Wimmer. 2023. "Influence Motives in Social Signaling: Evidence from COVID-19 Vaccinations in Germany." *American Economic Review: Insights*, 5(2): 275–291.

European Commission. 2022. *Cancer screening in the European Union.* Publications Office of the European Union.

Eyster, Erik, Matthew Rabin, and Georg Weizsacker. 2015. "An experiment on social mislearning." *SSRN 2704746 Working Paper.*

Forsythe, Robert, Russell Lundholm, and Thomas Rietz. 1999. "Cheap talk, fraud, and adverse selection in financial markets: Some experimental evidence." *Review of Financial Studies*, 12(3): 481–518.

Francetic, Igor, Rachel Meacock, and Matt Sutton. 2022. "Understanding Concordance in Health Behaviours among Couples: Evidence from the Bowel Cancer Screening Programme in England." *Journal of Economic Behavior & Organization*, 201: 310–345.

Frankel, Alex, and Navin Kartik. 2019. "Muddled information." *Journal of Political Economy*, 127(4): 1739–1776.

Fudenberg, Drew, and David K Levine. 1993. "Self-confirming equilibrium." *Econometrica*, 523–545.

Funk, Patricia. 2010. "Social Incentives and Voter Turnout: Evidence from the Swiss Mail Ballot System." *Journal of the European Economic Association*, 8(5): 1077–1103.

Galor, Oded, Ömer Özak, and Assaf Sarid. 2018. "Geographical Roots of the Coevolution of Cultural and Linguistic Traits." NBER Working Paper 25289.

Gentzkow, Matthew, and Jesse M Shapiro. 2011. "Ideological segregation online and offline." *Quarterly Journal of Economics*, 126(4): 1799–1839.

Gentzkow, Matthew, Michael B Wong, and Allen T Zhang. 2018. "Ideological bias and trust in information sources." *Unpublished manuscript.*

Gilovich, Thomas, and Kenneth Savitsky. 1999. "The spotlight effect and the illusion of transparency: Egocentric assessments of how we are seen by others." *Current Directions in Psychological Science*, 8(6): 165–168.

Gilovich, Thomas, Victoria Husted Medvec, and Kenneth Savitsky. 2000. "The spotlight effect in social judgment: an egocentric bias in estimates of the salience of one's own actions and appearance." *Journal of Personality and Social Psychology*, 78(2): 211.

Ginsburgh, Victor, and Shlomo Weber. 2020. "The Economics of Language." *Journal of Economic Literature*, 58(2): 348–404.

Goldberg, Jessica, Mario Macis, and Pradeep Chintagunta. 2022. "Incentivized Peer Referrals for Tuberculosis Screening: Evidence from India." *American Economic Journal: Applied Economics*, forthcoming.

Goldzahl, Léontine, Guillaume Hollard, and Florence Jusot. 2018. "Increasing breast-cancer screening uptake: A randomized controlled experiment." *Journal of Health Economics*, 58: 228–252.

Golman, Russell, David Hagmann, and George Loewenstein. 2017. "Information Avoidance." *Journal of Economic Literature*, 55(1): 96–135.

Golman, Russell, George Loewenstein, Andras Molnar, and Silvia Saccardo. 2022. "The Demand for, and Avoidance of, Information." *Management Science*, 68(9): 6454–6476.

Greiner, Ben. 2004. "An online recruitment system for economic experiments." *Forschung und wissenschaftliches Rechnen*, 63: 79–93.

Gulesci, Selim, Sam Jindani, Eliana La Ferrara, David Smerdon, Munshi Sulaiman, and H. Young. 2021. "A Stepping Stone Approach to Understanding Harmful Norms."

Guriev, Sergei, and Elias Papaioannou. 2022. "The Political Economy of Populism." *Journal of Economic Literature*, 60(3): 753–832.

Handel, Benjamin R., and Jonathan T. Kolstad. 2015. "Health Insurance for "Humans": Information Frictions, Plan Choice, and Consumer Welfare." *American Economic Review*, 105(8): 2449–2500.

Harrison, Glenn W., and John A. List. 2004. "Field Experiments." *Journal of Economic Literature*, 42(4): 1009–1055.

Hartzmark, Samuel M, Samuel D Hirshman, and Alex Imas. 2021. "Ownership, learning, and beliefs." *Quarterly Journal of Economics*, 136(3): 1665–1717.

Hauser, Christine. 2020. "Merriam-Webster Revises 'Racism' Entry After Missouri Woman Asks for Changes." *New York Times.* June 10. https://www.nytimes.com/2020/06/10/us/merriam-webster-racism-definition.html.

Hyde, Kelly. 2021. "Learning About Subjective Uncertainty: Overinference from Observable Characteristics in Disaggregated Data." Working Paper.

Intergovernmental Panel On Climate Change. 2022. *Climate Change and Land: IPCC Special Report on Climate Change, Desertification, Land Degradation, Sustainable Land Management, Food Security, and Greenhouse Gas Fluxes in Terrestrial Ecosystems.* Cambridge University Press.

Jewitt, Ian. 2004. "Notes on the 'Shape' of Distributions." Online Manuscript.

Jin, Ginger Zhe, Michael Luca, and Daniel Martin. 2021. "Is no news (perceived as) bad news? an experimental investigation of information disclosure." *American Economic Journal: Microeconomics*, 13(2): 141–73.

Juhn, Young J., Timothy J. Beebe, Dawn M. Finnie, Jeff Sloan, Philip H. Wheeler, Barbara Yawn, and Arthur R. Williams. 2011. "Development and Initial Testing of a New Socioeconomic Status Measure Based on Housing Data." *Journal of Urban Health*, 88(5): 933–944.

Kamenica, Emir, and Matthew Gentzkow. 2011. "Bayesian Persuasion." *American Economic Review*, 101(6): 2590–2615.

Karing, Anne. 2023. "Social Signaling and Childhood Immunization: A Field Experiment in Sierra Leone." Working Paper.

Katalinic, Alexander, Nora Eisemann, Klaus Kraywinkel, Maria R. Noftz, and Joachim Hübner. 2020. "Breast cancer incidence and mortality before and after implementation of the German mammography screening program." *International Journal of Cancer*, 147(3): 709–718.

Khan, Jahidur Rahman, Suzanne Jane Carroll, Matthew Warner-Smith, David Roder, and Mark Daniel. 2021. "Residential area and screening venue location features associated with spatial variation in breast cancer screening invitation response rates: an observational study in Greater Sydney, Australia." *BMJ Open*, 11(4): e043853.

Kolip, Petra, and Klaus-Dieter Wurche. 2005. "Mammografie Screening, Was MultiplikatorInnen vor Ort wissen sollten, Erfahrungen-Informationen-Tipps." Institut für Public Health und Pflegeforschung der Universität Bremen.

Kowalski, Amanda E. 2023. "Behaviour within a Clinical Trial and Implications for Mammography Guidelines." *Review of Economic Studies*, 90(1): 432–462.

Krupka, Erin L, and Roberto A Weber. 2013. "Identifying Social Norms Using Coordination Games: Why Does Dictator Game Sharing Vary?" *Journal of the European Economic Association*, 11(3): 495–524.

Lane, Tom, Daniele Nosenzo, and Silvia Sonderegger. 2023. "Law and Norms: Empirical Evidence." *American Economic Review*, 113(5): 1255–1293.

Lemke, Dorothea, Shoma Berkemeyer, Volkmar Mattauch, Oliver Heidinger, Edzer Pebesma, and Hans-Werner Hense. 2015. "Small-area spatio-temporal analyses of participation rates in the mammography screening program in the city of Dortmund (NW Germany)." *BMC Public Health*, 15(1): 1190.

Levy, Gilat, and Ronny Razin. 2019. "Echo chambers and their effects on economic and political outcomes." *Annual Review of Economics*, 11: 303–328.

Liao, Shen-Yi, and Nat Hansen. 2023. "'Extremely Racist' and 'Incredibly Sexist': An Empirical Response to the Charge of Conceptual Inflation." *Journal of the American Philosophical Association*, 9(1): 72–94.

Loewenstein, George, Joelle Y. Friedman, Barbara McGill, Sarah Ahmad, Suzanne Linck, Stacey Sinkula, John Beshears, James J. Choi, Jonathan Kolstad, David Laibson, Brigitte C. Madrian, John A. List, and Kevin G. Volpp. 2013. "Consumers' misunderstanding of health insurance." *Journal of Health Economics*, 32(5): 850–862.

Macchi, Elisa. 2023. "Worth Your Weight: Experimental Evidence on the Benefits of Obesity in Low-Income Countries." *American Economic Review*, Forthcoming.

Manski, Charles F. 1993. "Identification of Endogenous Social Effects: The Reflection Problem." *Review of Economic Studies*, 60(3): 531–542.

Marmaros, David, and Bruce Sacerdote. 2006. "How do friendships form?" *Quarterly Journal of Economics*, 121(1): 79–119.

McPherson, Miller, Lynn Smith-Lovin, and James M Cook. 2001. "Birds of a Feather: Homophily in Social Networks." *Annual Review of Sociology*, 27(1): 415–444.

Miles, Robert. 1989. "Conceptual inflation." In *Racism.* 41–68. Routledge.

Milkman, Katherine L., Mitesh S. Patel, Linnea Gandhi, Heather N. Graci, Dena M. Gromet, Hung Ho, Joseph S. Kay, Timothy W. Lee, Modupe Akinola, John Beshears, Jonathan E. Bogard, Alison Buttenheim, Christopher F. Chabris, Gretchen B. Chapman, James J. Choi, Hengchen Dai, Craig R. Fox, Amir Goren, Matthew D. Hilchey, Jillian Hmurovic, Leslie K. John, Dean Karlan, Melanie Kim, David Laibson, Cait Lamberton, Brigitte C. Madrian, Michelle N. Meyer, Maria Modanu, Jimin Nam, Todd Rogers, Renante Rondina, Silvia Saccardo, Maheen Shermohammed, Dilip Soman, Jehan Sparks, Caleb Warren, Megan Weber, Ron Berman, Chalanda N. Evans, Christopher K. Snider, Eli Tsukayama, Christophe Van den Bulte, Kevin G. Volpp, and Angela L. Duckworth. 2021. "A megastudy of text-based nudges encouraging patients to get vaccinated at an upcoming doctor's appointment." *Proceedings of the National Academy of Sciences*, 118(20): e2101165118.

Mokdad, Ali H., James S. Marks, Donna F. Stroup, and Julie L. Gerberding. 2004. "Actual causes of death in the United States, 2000." *Jama*, 291(10): 1238–1245.

Morris, Stephen. 2001. "Political correctness." *Journal of Political Economy*, 109(2): 231–265.

Oprea, Ryan, and Sevgi Yuksel. 2022. "Social exchange of motivated beliefs." *Journal of the European Economic Association*, 20(2): 667–699.

Oster, Emily, Ira Shoulson, and E. Ray Dorsey. 2013. "Optimal Expectations and Limited Medical Testing: Evidence from Huntington Disease." *American Economic Review*, 103(2): 804–830.

Peisakhin, Leonid, and Arturas Rozenas. 2018. "Electoral effects of biased media: Russian television in Ukraine." *American Journal of Political Science*, 62(3): 535–550.

Piketty, Thomas. 1995. "Social mobility and redistributive politics." *Quarterly Journal of Economics*, 110(3): 551–584.

Pruckner, Gerald J., Thomas Schober, and Katrin Zocher. 2020. "The company you keep: health behavior among work peers." *European Journal of Health Economics*, 21(2): 251–259.

Robitaille, Nicole, Julian House, and Nina Mazar. 2021. "Effectiveness of Planning Prompts on Organizations' Likelihood to File Their Overdue Taxes: A Multi-Wave Field Experiment." *Management Science*, 67(7): 4327–4340.

Ross, Lee, David Greene, and Pamela House. 1977. "The "false consensus effect": An egocentric bias in social perception and attribution processes." *Journal of Experimental Social Psychology*, 13(3): 279–301.

Salmon, Charlotte, Marie-Élise Parent, Amélie Quesnel-Vallée, and Tracie A. Barnett. 2022. “A scoping review of social relationships and prostate cancer screening.” *Preventive Medicine*, 154: 106892.

Schwardmann, Peter. 2019. “Motivated health risk denial and preventative health care investments.” *Journal of Health Economics*, 65: 78–92.

Smith, Dinah, Katie Thomson, Clare Bambra, and Adam Todd. 2019. “The breast cancer paradox: A systematic review of the association between area-level deprivation and breast cancer screening uptake in Europe.” *Cancer Epidemiology*, 60: 77–85.

Spence, Michael. 1973. “Job Market Signaling.” *Quarterly Journal of Economics*, 87(3): 354–374.

Strømsnes, Kristin. 2008. “The importance of church attendance and membership of religious voluntary organizations for the formation of social capital.” *Social Compass*, 55(4): 478–496.

Sunstein, Cass R. 1996. “On the Expressive Function of Law.” *University of Pennsylvania Law Review*, 144(5): 2021–2053.

Sutter, Matthias, Silvia Angerer, Daniela Glätzle-Rützler, and Philipp Lergetporer. 2015. “The Effect of Language on Economic Behavior: Experimental Evidence from Children’s Intertemporal Choices.” IZA DP No. 9383.

Tabellini, Guido. 2008. “*Presidential Address* Institutions and Culture.” *Journal of the European Economic Association*, 6(2-3): 255–294.

tagesschau. 2023. “‘Klimaterroristen’ ist laut Jury Unwort des Jahres 2022.”

The Economist. 2021. “How to talk about Xinjiang.” *The Economist.* February 13.

Tillmanns, Hanna, Gerhard Schillinger, and Hendrik Dräther. 2021. “Inanspruchnahme von Früherkennungsleistungen der gesetzlichen Krankenversicherung durch AOK-Versicherte im Erwachsenenalter. 2009-2020.” Wissenschaftliches Institut der AOK.

Torres, Carlos, Lucy Ogbu-Nwobodo, Marcella Alsan, Fatima Cody Stanford, Abhijit Banerjee, Emily Breza, Arun G. Chandrasekhar, Sarah Eichmeyer, Mohit Karnani, Tristan Loisel, Paul Goldsmith-Pinkham, Benjamin A. Olken, Pierre-Luc Vautrey, Erica Warner, Esther Duflo, and COVID-19 Working Group. 2021. “Effect of Physician-Delivered COVID-19 Public Health Messages and Messages Acknowledging Racial Inequity on Black and White Adults’ Knowledge, Beliefs, and Practices Related to COVID-19: A Randomized Clinical Trial.” *JAMA Network Open*, 4(7): e2117115.

Traunmüller, Richard. 2011. “Moral Communities? Religion as a Source of Social Trust in a Multilevel Analysis of 97 German Regions.” *European Sociological Review*, 27(3): 346–363.

Tversky, Amos, and Daniel Kahneman. 1973. “Availability: A heuristic for judging frequency and probability.” *Cognitive Psychology*, 5(2): 207–232.

Ware, Jordan K. 2019. “Property value as a proxy of socioeconomic status in education.” *Education and Urban Society*, 51(1): 99–119.

Wild, Christopher P., Elisabete Weiderpass, and Bernard W. Stewart. 2020. “World Cancer Report: Cancer Research for Cancer Prevention.” International Agency for Research on Cancer, Lyon, France.

Wittgenstein, Ludwig. 1953. *Philosophische Untersuchungen.* . 11 ed., Frankfurt am Main, Germany:Suhrkamp Verlag.

World Health Organisation. 2019. "WHO's Global Health Estimates (GHE)."

Wübker, Ansgar. 2014. "Explaining variations in breast cancer screening across European countries." *The European Journal of Health Economics*, 15(5): 497–514.

Zentrum für Krebsregisterdaten. 2021. "Altersstandardisierte Krebssterblichkeit ging auch 2020 weiter zurück."