Habitual communication[*]

Konstantinos Ioannidis[†]

September 17, 2024

Abstract

This paper studies habitual communication in a sender-receiver setting with information asymmetry. We investigate how habits formed in familiar environments affect communication in an unfamiliar environment. Using a controlled experiment with varying levels of preference alignment, we test two hypotheses: (i) whether familiarity with common-interest compared to conflicting-interest environments leads to more informative communication in an unfamiliar environment, and (ii) how reliance on communication habits varies based on the frequency of interacting in an unfamiliar environment. We find evidence for habitual communication only when the unfamiliar environment occurs rarely. Analysis of individual decisions provides suggestive evidence on the mechanisms.

Keywords: Habits, Strategic information transmission, Communication, Experiment
JEL Codes: D91, C92, D01, D83

1 Introduction

A wide range of everyday activities are repeated frequently. Repetition facilitates habit formation, and habits allow us to make decisions on auto-pilot. In other words, the answer to why people act the way they do is often simply because they are used to (Wood et al., 2002; Lally et al., 2010). This paper studies habitual behaviour in the context of strategic communication between an informed sender and an uninformed receiver. Many economically relevant interactions are characterised by such information asymmetry. Investors receive advice from financial advisors (Inderst and Ottaviani, 2012; Angelova and Regner, 2013), patients receive treatment information from physicians (Guo et al., 2020), and consumers receive product information from salespeople (Gardete, 2013; Chakraborty and Harbaugh, 2014). Habitual communication in such settings has important implications. A real estate agent who works in a seller's market, where demand exceeds supply, may develop the habit of negotiating hard as they have high bargaining power. Would they adapt their bargaining strategy in situations when supply increases and how does this depend on the steepness of the increase? An investor during prosperous times

[*] Financial support from the Research Priority Area Behavioral Economics of the University of Amsterdam, and the Leverhulme International Professorship in Neuroeconomics at the University of Cambridge are gratefully acknowledged. The author would like to thank Marta Serra-Garcia, Joel Sobel, and Joseph Tao-yi Wang for their valuable feedback on the paper as well as conference participants at IAREP/SABE (2024), EEA (2024), SEET (2024), BEEL (2022), CeDEx-CBESS-CREED meetings (2020, 2021), and ESA (2021). The author would like to thank Tina Marjanov for her help in programming the experiment. All data and code can be found on Github.

[†] University of Cambridge. Corresponding e-mail: ioannidis.a.konstantinos@gmail.com

may develop the habit of investing in high-risk high-return assets. Would they adjust their risk portfolio sufficiently when they only rarely receive signals that the economy is slowing down compared to an extensively media covered emerging crisis?

We focus on strategic communication in the form of strategic information transmission between two asymmetrically informed agents where (i) preferences are misaligned and (ii) messages do not directly affect monetary payoffs. In a seminal paper, Crawford and Sobel (1982) analysed such cheap talk games and showed that communication becomes less informative when the preferences of the sender and the receiver diverge. Standard economic theory leaves no scope for habits to affect the informativeness of communication. Primarily interacting in common-interest settings may facilitate the formation of habits of truth-telling and believing messages. Primarily interacting in conflicting-interest settings may facilitate the formation of habits of lying and distrusting messages. This paper provides empirical evidence for this line of reasoning.

We provide a behavioural model built on the assumption that with positive probability the agent does not adapt their strategy when the preference alignment changes. The model provides a simple framework to motivate our predictions. Specifically, we are interested to experimentally test two hypotheses: (i) whether familiarity with common-interest compared to conflicting-interest environments leads to more informative communication in an unfamiliar environment, and (ii) whether familiarity with common-interest environments leads to overcommunication whereas familiarity with conflicting-interest environments leads to undercommunication. We measure the informativeness of the communication by the correlation between states and actions.

We use a controlled laboratory experiment to address our research questions. Our participants play multiple rounds of a cheap talk sender-receiver game. In each round, the payoff-relevant state of the world is randomly drawn. The sender observes the true state whereas the receiver does not. The sender sends a message about the state to the receiver who chooses an action determining the payoffs of both players. Our treatments vary the preference alignment between the two players. We use a 2×2 between-participants treatment design. The participants play 60 rounds of the sender-receiver game with either fully conflicting, partially aligned or fully aligned interests. The 60 rounds are divided in two parts of 30 rounds each. The treatments vary (i) whether sender and receiver start with having conflicting or aligned interests in all 30 rounds of part one and (ii) whether they have partially aligned interests throughout all the remaining 30 rounds or rarely so (randomly in 10 out of 30 rounds). Our primary data are the choices of participants, i.e. sender messages and receiver actions. Additionally, we record decision times, cognitive ability (via the CRT), and risk attitudes and trust attitudes.

Part one facilitates the formation of different communication habits. We use 30 rounds as habit formation requires long repetition in a stable environment (Wood and Rünger, 2016). We are interested in the effect of the (potentially) formed habits on communication in the unfamiliar environment with partially aligned preferences. We hypothesise that communication will be more informative for participants who started with the aligned environment than for participants who started with the conflicting environment. Part two varies how often participants interact in the unfamiliar environment. Motivated by the psychology and neuroscience literature, we conjecture that reliance on habits will be stronger when the new environment occurs rarely.

Our main finding is that communication under partially aligned interests is more informative for participants who started with common interests in part one, but only if they interact in the new environment rarely. This effect persists over time. When the new environment occurs frequently, participants quickly adapt their behaviour and we find no effect of part one in the informativeness of communication. Additionally, the observed correlations provide point estimates of the informativeness of communication. We find that, compared to the most informative perfect Bayesian equilibrium, participants who started with the common-interest environment overcommunicate whereas participants who started with the conflicting-interest environment undercommunicate.

Our design and the behavioural model are not restricted to a single mechanism for why people do not always adapt their strategy when facing an unfamiliar environment. To better understand the mechanisms at the individual level, we classify participants as habitual if their choices satisfy two conditions: (i) they use a stable strategy for the majority of decisions in part one, and (ii) they use the same strategy when interacting in the new environment. This exercise reveals that multiple mechanisms play a role in habitual communication. Some participants do not even notice the change in the environment whereas some participants do but consciously decide to stick with their strategy. Expectedly, habitual participants make decisions faster and have lower CRT scores.

Our paper speaks to various strands of research. First, it is part of the economic literature on habit formation. Many studies focus on consumption habits, and, more specifically, on the effect of past consumption on future consumption (see Havranek et al. (2017) for a literature review and meta analysis of relevant studies). A characteristic example of habits in consumption comes from Camerer et al. (2024) who document that consumers do not disrupt their habitual purchasing of tuna cans, despite a change in the size of the can. Empirical evidence document habits in a range of settings such as saving (de Mel et al., 2013; Schaner, 2018), exercising (Charness and Gneezy, 2009; Acland and Levy, 2015; Royer et al., 2015; Buyalskaya et al., 2023), voting (Gerber et al., 2003; Meredith et al., 2009; Coppock and Green, 2016; Fujiwara et al., 2016), usage of public transport (Gravert and Collentine, 2021), water consumption (Byrne et al., 2023), electricity usage (Ito et al., 2018), social media posting (Camerer et al., 2024), and hand washing (Hussam et al., 2022; Buyalskaya et al., 2023).[1] We contribute to this literature by providing experimental evidence of habits in a strategic setting, and more specifically in strategic communication, whereas the literature on habits typically focuses on individual decisions.

Closely related is literature studying spillover effects. Experiments have shown that establishing higher prosociality with one task results in higher cooperation in a different task, both when participation in the two tasks is sequential (Knez and Camerer, 2000; Cassar et al., 2014; Peysakhovich and Rand, 2016; Stagnaro et al., 2017; Duffy and Fehr, 2018) and when it is simultaneous (Falk et al., 2013; McCarter et al., 2014). Engl et al. (2021) study whether behavioural spillovers operate through preferences or beliefs in sequential and in simultaneous decision making. Our research provides habits as an alternative mechanism for creating prece-

[1] This literature is quite heterogeneous in their operationalisation of habits. For example, voting habits have been documented as the excess probability of voting after having voting once in the past election. This notion of habit may not satisfy the notion of habit from psychology and neuroscience, which is also the view we share in this paper, as that view requires a high degree of repetition.

dents. It remains an open question on whether precedents acquired by habit formation, which need long repetition, or acquired by different institutions or simultaneous decisions in a different task, create behavioural spillovers through different mechanisms.

Second, our paper belongs in the line of experimental cheap talk games. Starting from Dickhaut et al. (1995), a long list of experiments have investigated the comparative statics of Crawford and Sobel (1982). A common finding is overcommunication; participants typically communicate more information than the most informative equilibrium predicted by theory (Cai and Wang, 2006; Sánchez-Pagés and Vorsatz, 2007; Kawagoe and Takizawa, 2009; Wang et al., 2010; de Haan et al., 2015).[2] The causes of overcommunication have been debated with Lafky et al. (2022) finding that it is driven by heterogeneity in strategic reasoning instead of preferences, and Li et al. (2022) finding that it is driven by trust instead of strategic reasoning. Our design allows us to test the conjecture that overcommunication is observed because participants are used to common-interest environments outside of the lab, where habits of honest informative communication may form. When participating in an experiment, participants may carry this predisposition towards honest communication with them. By varying the environment in which habits are formed, we observe both overcommunication and undercommunication, which is consistent with our conjecture. Thus, our results offer a novel alternative explanation for the origins of overcommunication.

A closely related research question is explored in Belot and van de Ven (2019). They expose participants to either low and high incentives to lie in a sender-receiver game and reverse the incentives halfway through the experiment. They find no evidence of persistency of either honest or dishonest communication. However, in their experiment participants played 12-14 rounds in total whereas in ours they played 60 rounds. As also mentioned in their discussion, habit formation takes time and their shorter experiment may not have been able to facilitate it.

Third, our results speak to the literature documenting communication differences between individuals (Sánchez-Pagés and Vorsatz, 2007; Hurkens and Kartik, 2009; Serota et al., 2010) as well as between groups such as countries (Holm and Kawagoe, 2010; Innes and Arnab, 2013; Pascual-Ezama et al., 2015; Hugh-Jones, 2016), occupation (Cohn et al., 2014, 2015) and religiosity (Arbel et al., 2014). With a randomised controlled experiment, we present evidence for a causal link between habit formation in a familiar environment and communication in an unfamiliar environment. While models allowing for heterogeneous lying aversion can incorporate such differences, our results suggest that habits can also explain those differences without the need for heterogeneous preferences.

The remaining of the paper is organised as follows. Section 2 provides a detailed presentation of the sender-receiver game and the experimental design. There we also provide a behavioural model to derive our predictions. All results are presented in Section 3. We end the paper with Section 4 which discusses the results, positions the contributions, and suggests areas for future research.

[2] A comprehensive literature review of experimental cheap talk games can be found in Blume et al. (2020).

2 Design & Predictions

2.1 The sender-receiver game

The experiment considers a discrete cheap talk game between one sender and one receiver. In the beginning of each round, the state of the world (s) is uniformly drawn from the set $S = \{1, 2, 3, 4, 5\}$. The prior distribution is commonly known. The sender privately observes the draw and has to send a message (m) to the receiver. The possible messages are of the form "*The state is m*", where $m \in M = \{1, 2, 3, 4, 5\}$. The receiver is uninformed about the true state of the world. After observing the sender's message, the receiver chooses an action (a) from the set $A = \{1, 2, 3, 4, 5\}$. The action determines the payoffs of both players and ends the round.

The payoffs for both players depend only on the state and the action (and not on the message), and are given below.[3]

$$U^S(a, s, b) = 110 - 20|s + b - a|^{1.4} \text{ and } U^R(a, s) = 110 - 20|s - a|^{1.4}$$

From the (induced) utility functions, it is clear that the receiver optimally wants to match the true state ($a = s$) whereas the sender wants the receiver to choose an action higher than the state ($a = s + b$). Thus, the parameter b naturally captures the alignment of interests between the sender and the receiver; the larger the bias, the more misaligned their preferences.

2.2 Treatments

The participants play 60 rounds of the sender-receiver game. The rounds are split in part one (rounds 1-30) and part two (rounds 31-60). We use a 2×2 between participants design varying the value of the bias parameter in the two parts. We emphasise that the participants are only aware that they will play 60 rounds, but not that there are two parts.

Part one is either *Aligned* or *Conflict.* In Aligned, the participants play 30 rounds with a bias parameter of $b = 0.2$. In Conflict, the participants play 30 rounds with a bias parameter of $b = 2$. Part two is either *Rare* or *Frequent.* In Frequent, the participants play all rounds of part two with a bias parameter of $b = 1$. In Rare, the participants play 10 rounds with $b = 1$ and 20 rounds with the same bias parameter as in part one ($b = 0.2$ if they started with Aligned and $b = 2$ if they started with Conflict). The rounds with $b = 1$ are randomly chosen, but they drawn beforehand and are kept constant across all the Rare sessions. Overall, our design has four treatments, namely *Aligned-Rare, Aligned-Frequent, Conflict-Rare, Conflict-Frequent.* They are visualised in Figure 1.

As can be seen from the figure, in each round the bias is slightly perturbed. The noise is small enough that the overall incentive structure is not affected. This design choice aimed at minimising experimental demand effect when the underlying bias changes. Without the noise, the bias would change after being the same for 30 rounds. This could alert participants into

[3] The payoff functions are taken from Cai and Wang (2006) and Wang et al. (2010). The value of 1.4 in the exponent is used to enhance payoff differences across receiver actions. Cai and Wang (2006) used various values as a robustness check with similar results.

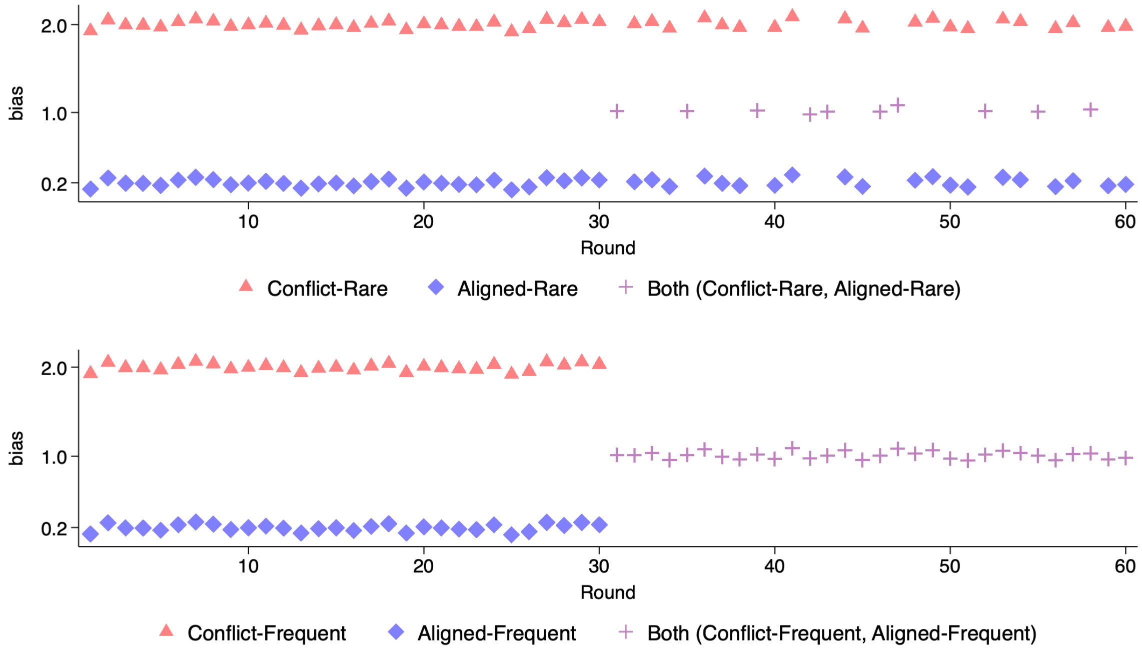


Figure 1: Bias per round for each treatment

thinking they should behave differently. With the small noise, their payoffs slightly change in every round, and more sharply change when the underlying bias also changes.

2.3 Behavioural model

We provide a behavioural model to derive our experimental predictions. As a benchmark, we assume that the agent chooses a strategy according to a perfect Bayesian equilibrium. The sender optimally chooses a message given the state and the bias, anticipating the optimal decision of the receiver. We denote the sender's best response function as $BR^S(s, b) \in M$. The receiver, after seeing the message and updating their beliefs about the state, chooses an action. The bias naturally enters their belief updating. We denote the receiver's best response function as $BR^R(m, b) \in A$. An equilibrium is a combination of mutual best responses.

Crawford and Sobel (1982) analysed such games and showed that all equilibria are partition equilibria. In such an equilibrium, the sender partitions the state space and randomly selects one message from each element of the partition. The message reveals which element of the partition the true state belongs to. The larger the bias parameter, the more coarse the partition is. In other words, less information is revealed by the sender and less faith is placed in the message by the receiver when their preferences are less aligned. Typically there exist multiple equilibria for each value of b. Crawford and Sobel (1982) showed that the most informative equilibrium is Pareto superior to all other equilibria.[4]

In our treatments, the bias changes in part two. We denote the current bias as b and

[4] While some equilibrium selection criteria are too strict and eliminate all equilibria in Crawford-Sobel like games (Matthews et al., 1991; Farrell, 1993), criteria that do select an equilibrium, typically also select the most informative one (Chen et al., 2008; de Groot Ruiz et al., 2015).

the bias of part one as b^*. Our behavioural assumption is that with positive probability the agent does not adapt their strategy when the bias changes. In this case, they choose their best response according to b^* instead of b. We make no assumption about the mechanism driving the inadaptability of behaviour. The agents may fail to adapt their strategy because they do not correctly observe the current bias, an assumption which resembles the salience literature (Bordalo et al., 2022).[5] The agents may also stick with their strategy from part one, despite correctly observing the current bias, in order to economise on the cognitive costs of reoptimising, an assumption which resembles the rational inattention literature (Sims, 2003; Maćkowiak et al., 2023) or because they hold the belief that their strategic counterpart will not adapt their strategy. Our model incorporates all those mechanisms in a reduced form.

We denote the probability that the agent does not adapt their strategy as w, and define the behavioural best response as the linear mixture of the perfect Bayesian best responses that correspond to the current and the previous bias.[6] Formally, the behavioural best responses for the sender and the receiver are denoted as:

$$BBR^S(s, b, b^*, w) = w \cdot BR^S(s, b) + (1 - w)BR^S(s, b^*)$$
$$BBR^R(m, b, b^*, w) = w \cdot BR^R(m, b) + (1 - w)BR^R(m, b^*)$$

When $w = 1$ the agents fully adapt their strategy, so the bias from part one does not influence their behaviour. In that case, the model collapses to the standard Crawford-Sobel setting, and agents behave according to the perfect Bayesian equilibrium. On the other extreme, when $w = 0$, they never adapt their strategy. In that case, they again behave according to the perfect Bayesian equilibrium but under the wrong bias value. Thus, their behaviour is entirely driven by the bias from part one. For intermediate values, they sometimes behave according to the current bias and sometimes according to the bias they are familiar with. Our treatments vary the probability of being in the unfamiliar environment, and we assume that the probability of reacting to the new bias is lower in Rare than in Frequent. Thus, behaviour depends more strongly on the past environment when the unfamiliar environment occurs less often.

Our reduced-form model captures the intuition behind prominent models in psychology, computational neuroscience, and neurobiology. Psychology models of habitual behaviour assume that habits are less likely to persist if the chain of stimuli-action-outcome is disturbed. In our setup, when the bias changes, the same action leads to a different payoff. Thus, while the stimuli is fixed, the action-outcome correspondence is disturbed, and more frequently so in Rare treatments. Similarly, computational neuroscience models such as reinforcement learning and reference-model based learning (also known as model-reference adaptive control) assume that agents change behaviour and increase their learning rate when the expected prediction error of an action-outcome correspondence is large.[7] In our setup, the difference in payoffs resulting for

[5] Bordalo et al. (2022) distinguish three distinct mechanisms which can drive bottom-up attention and salience: (i) contrast, defined as the distance of the value of an attribute of a good from the average value of that attribute in the consideration set of goods, (ii) surprise, defined as the distance of attribute values from average values of that attributed that are recalled from memory, and (iii) prominence, defined as the degree that an attribute is easily observed. Our behavioural assumption is closer to the latter case.

[6] Our formulation of the behavioural best response as a weighted average shares structure with behavioural inattention (Gabaix, 2019) and behavioural attenuation (Enke et al., 2024).

[7] Condorelli and Furlan (2023) apply a reinforcement learning approach to the Crawford and Sobel (1982)

the same action when the bias has changed is fixed between treatments, but the frequency of observing this difference is larger in Rare. Thus, while the magnitude of the prediction error is fixed, it is larger in expectation in Rare due to the higher frequency of observing it. Furthermore, the current view in neurobiology is that humans facing the trade-off between specialisation in a given environment and robustness/adaptability in a new environment (Amaya and Smith, 2018). To solve this trade-off efficiently, they code more decision-relevant information which they expect to need more often (Glimcher, 2022). [8]

2.4 Predictions

Before presenting the predictions from the behavioural model, we walk the reader through the benchmark case of perfect Bayesian equilibrium. In our treatments, we use three bias values: $b = 0.2$, $b = 1$, and $b = 2$.[9]

When $b = 0.2$, the most informative equilibrium is a separating equilibrium. Since, the sender has no incentive to deceive the receiver, they send a truthful message. The receiver, knowing that the sender's message is perfectly revealing of the state, chooses an action that matches the message. When $b = 1$, the alignment of incentives is mild and the most informative equilibrium is a semi-separating equilibrium. The sender partitions the state space in two sets, namely $\{1\}$ and $\{2, 3, 4, 5\}$. The sender sends a truthful message when $s = 1$, and randomly selects a message for all other states. The receiver, chooses $a = 1$ if the message is "The state is 1" as in this case the message perfectly reveals the state. However, when seeing any other message, they only infer that the state is in $\{2, 3, 4, 5\}$, and so randomly pick between action 3 and action 4 with equal probability. When $b = 2$, the preference misalignment between the sender and the receiver is so large that only a babbling equilibrium exists. The sender randomly sends a message with equal probability. The receiver, realising the message is completely uninformative about the state, ignores the message and selects the ex-ante optimal action ($a = 3$).

All perfect Bayesian equilibria are presented in Table A1. The table is augmented with the correlation between state and action in each equilibrium. We use the correlation as our measure of the informativeness of communication (henceforth just correlation).[10] The correlation ranges from 0 to 1 corresponding to fully uninformative communication and fully informative communication respectively. In the aligned environment ($b = 0.2$) the most informative equilibrium is a separating equilibrium with $\rho = 1.00$. In the conflict environment ($b = 2$) the most informative equilibrium is a babbling equilibrium with $\rho = 0.00$. We are interested in behaviour when $b = 1$ where most informative equilibrium is a semi-separating equilibrium with $\rho = 0.65$.

We then solve the behavioural model for all values of w. We compute the correlation of the behavioural equilibrium from the behavioural best responses of the sender and the receiver for

setting and show that communication typically converges to the most informative perfect Bayesian equilibrium.

[8] In financial applications, efficient coding gives rise to phenomena such as outlier blindness (Payzan-LeNestour and Woodford, 2022).

[9] In the preregistration we had specified $b = 0$ instead of $b = 0.2$ to be the lowest bias value. We changed it to avoid making the coordination between the sender telling the truth and the receiver following the message too obvious. The underlying structure of the equilibria remains unchanged as can be seen in subsection A.1, where we present the full list of equilibria for all positive values of b.

[10] We choose the correlation between states and actions as a measure of informativeness to facilitate comparisons with previous experimental literature (Cai and Wang, 2006; Kawagoe and Takizawa, 2009; Wang et al., 2010).

all values of w. Figure 2 illustrates the informativeness of communication for all values of w. The (top) blue line corresponds to the case where part one is aligned, and the (bottom) red line corresponds to the case where part one is conflicting.

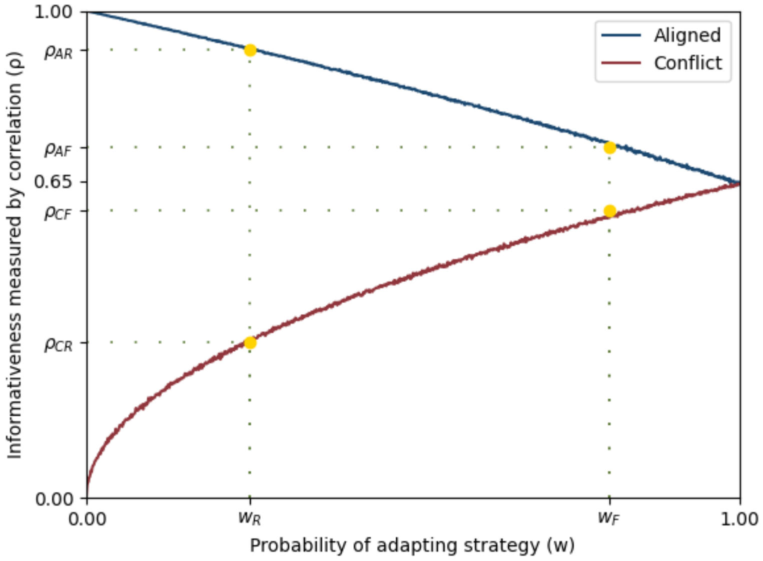


Figure 2: Informativeness of behavioural equilibria when $b = 1$

We assume that the probability of changing their strategy differs is higher in frequent than in rare treatments since the underlying bias changes more rarely in the latter. We note that, since the bias is slightly perturbed in each round, the probability of observing the change in the environment may be less than 1 even in the Frequent treatments. A key prediction of the behavioural model is that when the new environment occurs more rarely, the informativeness of communication is closer to the informativeness of part one. We illustrate this in Figure 2. We select a low and a high value of w, corresponding to the Rare (w_R) and Frequent treatments (w_F). The correlations for those values are indicated with dashed lines. We collect the results in the proposition below.

Proposition 1. $\rho_{AR} > \rho_{AF} > 0.65 > \rho_{CF} > \rho_{CR}$.

We first ask whether participants communicate habitually, separately for rare and frequent treatments. As can be seen from Proposition 1, we expect the likelihood of finding evidence supporting the former to be higher than for the latter. Formally, we test the following predictions.

Prediction 1. *Habitual communication in Rare:* $\rho_{AR} > \rho_{CR}$.

Prediction 2. *Habitual communication in Frequent:* $\rho_{AF} > \rho_{CF}$.

We then ask whether participating in a common-interest environment in part one leads to overcommunication, and participating in a conflicting-interest environment in part one leads to undercommunication.

Prediction 3. *(a) Overcommunication in Aligned-Rare:* $\rho_{AR} > 0.65$.

(b) *Undercommunication in Conflict-Rare:* $\rho_{CR} < 0.65$.

(c) *Overcommunication in Aligned-Frequent:* $\rho_{AF} > 0.65$.

(d) *Undercommunication in Conflict-Frequent:* $\rho_{CF} < 0.65$.

Given that in the rare treatments, the new environment occurs only in ten rounds, we split those ten rounds in the early rounds (first five) and late rounds (next five). To keep the amount of rounds interacting in the new environment –and consequently learning opportunities– fixed, we also split the first ten rounds in the frequent treatments and split them in early and late rounds accordingly. All predictions are tested separately for early and for late rounds to allow us to understand the persistence of habits.[11]

2.5 Procedures

The computerised laboratory experiment was conducted in October and November of 2020. All participants were recruited from the participant pool of the CREED laboratory of the University of Amsterdam. The experiment was programmed in oTree (Chen et al., 2016) and preregistered (Ioannidis, 2020). Each treatment arm had 64 participants, resulting in 256 participants in total. Our participants were on average 22 years old (mean = 22.37, sd = 4.29, min = 18, max = 60), primarily Economics students (64%), and evenly balanced across genders (52% female, 47% male, 1% other). Each participant only joined one session. They earned on average €27 (mean = 27.32, sd = 6.13, min = 6.95, max = 35.5) in approximately two hours.

The sessions were run during the Covid-19 pandemic when access to the physical lab was restricted. Thus, the experiment shares featured of both lab and online experiments. The structure and duration of the experiment as well as the participant pool are typical of lab experiments. The sessions ran on an online server with participants accessing the experimental platform remotely, and anonymised Zoom session parallel to each session to allow the experimenter to monitor any questions or technical issues from participants; features typical of online experiments. We consider our setting as close to a lab experiment as possible given the pandemic restrictions, and refer to it as a lab experiment.[12]

Each session consisted of 16 participants randomised into two matching groups of eight.[13] Each matching group was randomly assigned to a treatment. Within a matching group, the

[11] For Rare treatments, the early rounds are 31, 35, 39, 42, and 43, and the late rounds are 46, 47, 52, 55, and 58. For the Frequent treatments, the early rounds are 31-35 and the late rounds are 36-40. The results are qualitatively similar if we define the same rounds as early and late in both treatments. We also obtain the same conclusions when extending the late rounds in the Frequent treatments to be the last rounds (56-60).

[12] Given that the experiment was run remotely, connectivity issues could temporarily prevent participants from accessing the experiment. To avoid delaying the session, a maximum of 180 seconds was allowed per decision. The timer was initially hidden from participants and only appeared when they had 30 seconds left. The timer was shown in 32 out of 15,360 decision screens and in 236 out of 15,360 feedback screens. If a participant failed to make a decision within 180 seconds, they were flagged as inactive. This automatically resulted in 0 points for them in that round. Their partner received 100 points and was informed that their partner was inactive in that round. To ensure the session proceeded without further delays, the maximum time available was reduced by 30 seconds for every round a participant was inactive. Thus, if a participant was inactive for more than five consecutive rounds, they would be removed from the rest of the experiment. No participant was removed from any session. In total four senders and ten receivers (not paired with each other) were inactive for one round, and two receivers were inactive for two rounds. Thus, we later remove 18 observations from our analysis.

[13] Due to attendance issues, two sessions had only 14 participants and two sessions had 18 participants. Thus, two matching groups have less participants (six) and two matching groups have more participants (ten).

participants were randomly assigned a role (i.e. sender or receiver) and kept it throughout the experiment. To avoid framing, in the experiment players were referred to as player A (sender) and player B (receiver). They were informed that the main experiment will last 60 rounds and that their cumulative earnings from all rounds will be converted to euros at a ratio of 200:1. After reading the rules of the sender-receiver game, they had to correctly answer a series of understanding questions.

In the main experiment, they played 60 rounds of the sender-receiver game. They were randomly rematched within their matching group in every round to avoid reputation effects. Eight independent sequences of true states were drawn before the experiment and used for each matching group respectively. The same sequences were used for all treatments to eliminate any difference in the variation of true states across all treatments. To ensure that Aligned-Rare and Conflict-Rare treatments are as comparable as possible, we fixed the rounds in which $b = 1$ across all matching groups.

The payoffs for both players were shown in a table whenever they made their decisions; both when the senders were choosing a message, and when the receivers were choosing an action. At the end of each round, both players received complete feedback about the true state, the message sent, the action chosen, and the realised payoffs of both players. The feedback screen also included the payoff table, allowing participants to reflect on their decisions.

The experiment ended with three post-experiment questionnaires measuring risk attitudes, cognitive ability, and trust attitudes, as well as a survey of standard demographics (age, gender, field of study).

The first questionnaire measured risk attitudes using the lottery method of Eckel and Grossman (2002). The participants had to choose from a series of lotteries whose expected payoff increases with variance. Their decision was incentivised and realised by the computer. Given the informational asymmetry of the interaction, controlling for risk is necessary as, for example, risk averse receivers may choose the ex-ante optimal action ($a = 3$).

The second questionnaire measured cognitive ability using the Cognitive Reflection Test (CRT) of Frederick (2005). CRT consists of questions with intuitive, but wrong, answers and measures the tendency to override intuition and deliberately reflect on the correct answer. To avoid participants being familiar with the questions from previous experiments, we used a modified set of questions (Shenhav et al., 2012). Measuring cognitive ability is interesting as participants with lower CRT may over-rely on habits, thus adapting their behaviour less in the rounds where they play the unfamiliar game ($b = 1$). The CRT was also incentivised.

The third questionnaire measured general trust attitudes towards strangers. We used two questions adapted from the World Values Survey (Glaeser et al., 2000), namely: (i) "When I communicate with strangers, I tell them the truth.", and (ii) "When I communicate with strangers, they tell me the truth". We used a five-point Likert scale from -2 (strongly disagree) to +2 (strongly agree). Their attitudes were elicited to serve as a proxy for their baseline tendency towards honest communication. All else being equal, participants who are more trusting towards strangers outside the lab may have a higher chance of sending a truthful message as senders or following a message as receivers.

Finally, decision times were recorded throughout the whole experiment.

3 Results

All reported tests are two-sided. All analyses are done using data aggregated over participants in a matching group and over rounds, to ensure all comparisons use independent observations. This approach leaves us with eight independent observations per treatment. The upside is that differences which are significant with this conservative approach indicate very high confidence in the treatment effect. The downside is that some comparisons may be underpowered. To address the possible low power issue, in the appendix we show that all results presented here remain valid when estimating them econometrically.

Before proceeding with the primary analysis, we briefly comment on behaviour in part one. A successful manipulation would require two conditions: (i) communication to be more informative in Aligned compared to Conflict, and (ii) decision times being faster over time. The average correlation in the Aligned environment ($r_A = 0.953$) is significantly higher than the correlation in the Conflict environment ($r_C = 0.387$). Our participants on average make decisions 0.41 seconds faster in each round. In subsection A.2 we provide additional details on the behaviour in part one. Since the bias is fixed during part one, we also compare our results with previous experimental literature and show that past findings replicate.

3.1 Habitual communication at the aggregate level

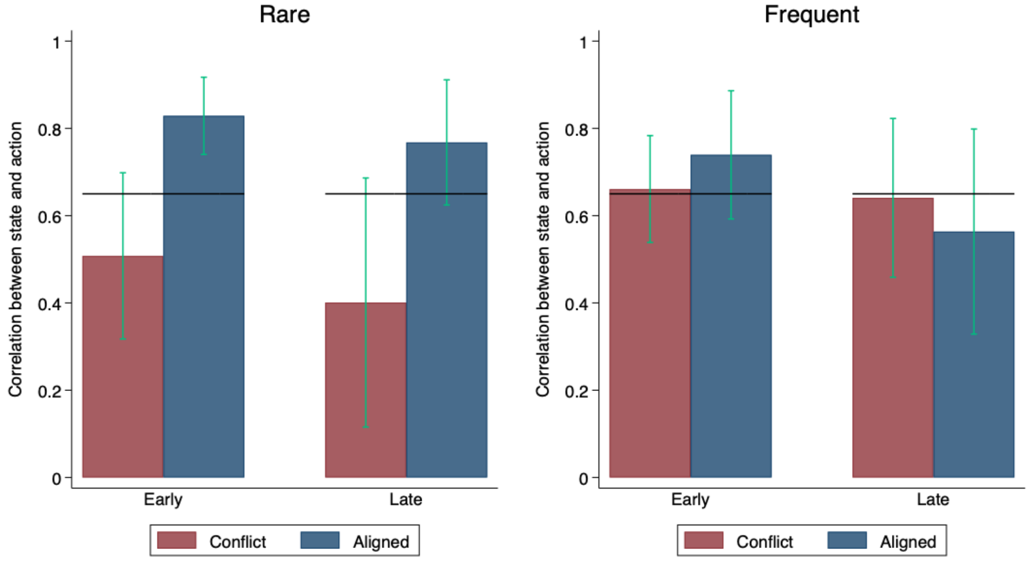
In this section we test our experimental predictions. We are interested to see whether our participants communicate habitually. Specifically, we test whether communication in the new environment is more informative for participants who participated in the Aligned compared to the Conflict environment in part one. We answer this question separately for Rare and Frequent treatments, allowing us to understand how the frequency of communicating in the unfamiliar environment moderates the tendency to communicate habitually. We also repeat the tests for early and late rounds, allowing us to investigate the persistence of habitual communication.[14]

We begin our analysis with Prediction 1. In the early rounds, the correlation in Aligned-Rare is significantly higher than the correlation in the Conflict-Rare ($r_{AR} = 0.829, r_{CR} = 0.508$, Wilcoxon ranksum test, $z = 2.731, p = 0.0047, N = 16$). The effect remains sizeable and significant in the late rounds ($r_{AR} = 0.768, r_{CR} = 0.401$, Wilcoxon ranksum test, $z = 2.310, p = 0.0207, N = 16$). The results are illustrated in the left part of Figure 3.

Result 1. *Communication in early rounds is more informative in Aligned-Rare than in Conflict-Rare treatment. The effect persists over time.*

We continue with Prediction 2. In the early rounds, the correlation in Aligned-Frequent and the correlation in the Conflict-Frequent are not significantly different ($r_{AF} = 0.734, r_{CF} = 0.661$, Wilcoxon ranksum test, $z = 0.945, p = 0.3823, N = 16$). In the late rounds, the correlation in Aligned-Frequent treatment and the correlation in the Conflict-Frequent treatment are also not significantly different ($r_{AF} = 0.564, r_{CF} = 0.641$, Wilcoxon ranksum test, $z = 0.525, p = 0.6454, N = 16$). The null effect is illustrated in the right part of Figure 3.

[14] In subsection A.3 we estimate ordered logistic regressions of receiver action on state, while clustering errors at the matching group level. Result 1 and Result 2 remain valid when estimated econometrically.



Lines in each bar indicate 95% confidence intervals. Horizontal lines indicate perfect Bayesian equilibrium correlation (0.650).

Figure 3: Treatment effects by frequency of the new environment

Result 2. *There is no difference in the informativeness of communication between Aligned-Frequent and Conflict-Frequent treatments, neither in early nor in late rounds.*

3.2 Overcommunication and undercommunication

We now turn our attention to the absolute levels of the correlations and test for overcommunication and undercommunication. Prediction 3 suggests that the observed correlations will be higher than 0.650 after Aligned and lower than 0.650 after Conflict. The results are visualised in Figure 3, where in each bar graph the horizontal black lines correspond to a correlation of 0.650.

Each comparison is performed by a signtest. In early rounds of the Aligned-Rare treatment, the correlation is higher ($r_{AR} = 0.829$, signtest, $p = 0.0039, N = 8$) whereas for Conflict-Rare the correlation is lower ($r_{CR} = 0.508$, signtest, $p = 0.1445, N = 8$) than 0.650. The same pattern is observed in late rounds ($r_{AR} = 0.768$, signtest, $p = 0.1445, N = 8$, $r_{CR} = 0.401$, signtest, $p = 0.0352, N = 8$). All tests find no evidence that the correlation differs from 0.650 (p-values are between 0.363 and 0.634) in either early or late of Aligned-Frequent and Conflict-Frequent treatments.[15]

Result 3. *Overcommunication is observed in Aligned-Rare treatment and undercommunication in Conflict-Rare. The informativeness of communication in the Aligned-Frequent and Conflict-Frequent treatments does not differ from perfect Bayesian equilibrium.*

[15] We emphasise that with eight observations, the signtest is conventionally significant if at least seven observations are in the predicted direction. Thus, a p-value of 0.1445 implies that the observed effect was in the predicted direction for six out of eight matching groups. To address the low power of the signtest, subsection A.4 presents results from the regression method suggested by Cai and Wang (2006). All conclusions remain valid with this alternative high-powered method.

3.3 Habitual communication at the individual level

The results presented so far are based on aggregate data. In this subsection, we look more closely at individual decisions to better understand habitual communication. Our primary goal is to provide qualitative evidence supporting our primary treatment effects. Thus, we want to verify whether more participants rely on communication habits when the new environment occurs rarely compared to frequently. Next, we provide evidence on the prevalence of the mechanisms behind reliance on communication habits. Finally, we perform a series of comparisons between habitual and non-habitual participants providing the reader with context of how their behaviours differ. More specifically, we test: (i) whether habitual participants make decisions faster, (ii) whether habitual participants show have lower cognitive ability, and (iii) whether the simplicity of the Aligned environment results in the formation of stronger communication habits than the more complex Conflict environment.

We first need a procedure to classify participants into habitual and non-habitual. We classify a participant as habitual if their decisions satisfy two requirements, taken directly from the psychology definition of habits. Habits are characterised by (i) high automaticity, and (ii) reduced dependence on goals (Wood and Rünger, 2016). We operationalise those two conditions as follows. For high automaticity, we require participants to converge to a stable strategy in part one. Since the habit formation process takes time (Lally et al., 2010), we ignore the first ten rounds where participants could potentially still be using trial and error. For reduced dependence on goals, we require participants to use the same stable strategy in part two as they did in part one, despite the change in the underlying bias. A participant is classified as *habitual* if their decisions satisfy *both* requirements. In simpler terms, we require participants to use a stable strategy in part one, and use the *same* stable strategy in part two.

Next, we need to define the set of strategies to consider. We do not restrict ourselves to a particular theoretical model. Instead, we take a data-driven approach and consider all pure strategies that can exist in the game. For each of the five states observed, senders can choose among five messages, resulting in 3,125 possible strategies. Symmetrically, for each of the five messages received, the receivers can choose among five actions, also resulting in 3,125 possible strategies.[16] For each strategy, we compute the percentage of decisions consistent with it. The consideration set consists of strategies which are consistent with at least 60% of participant decisions. If the set consists of more than one strategies, we select the one which matches the highest percentage of decisions. The threshold of 60% is used for both part one and part two.

With this procedure, we can successfully identify behavioural strategies for 235 participants for part one and for 242 participants for part two. In total, 112 participants use the same behavioural strategy in part one and part one, and consequently are classified as habitual.

[16] In similar experiments, individual decision analysis typically focused on level-k classification of behavioural types (Cai and Wang, 2006; Wang et al., 2010). With our procedure, additional strategies are also included. For example, when $b = 2$, no level-k prediction would imply that senders should exaggerate the true state by one. L0 senders would tell the truth, L1 senders should exaggerate by two since they believe they are facing credulous receivers, and higher level senders would exaggerate even more. Other econometric methods to estimate behavioural strategies are the Structural Frequency Estimation Method of Dal Bó and Fréchette (2011) and the spike-logit model of Costa-Gomes and Crawford (2006). In those methods, the set of candidate strategies is predefined. Costa-Gomes and Crawford (2006) consider whether alternative strategies (pseudotypes) provide a better fit than the original strategies as a robustness check for their classifications. Our method has a similar intuition in the sense that we consider every possible strategy and choose the best fitting one.

Table 1 shows the number of habitual participants across treatments, separated by their role.[17]

Role	Treatment				Total
	A-F	A-R	C-F	C-R	
Sender	14	11	10	13	48
Receiver	16	25	11	12	64
Total	30	36	21	25	112

Treatment abbreviations:
A-F = Aligned-Frequent
A-R = Aligned-Rare
C-F = Conflict-Frequent
C-R = Conflict-Rare

Table 1: Habitual participants per treatment

First, we find suggestive evidence that more participants communicate habitually when they face the new environment rarely compared to frequently. Taken together, in Aligned-Rare and Conflict-Rare 61 participants communicate habitually compared to 51 in Conflict-Frequent and Conflict-Rare. Our data is in the expected direction (proportion test, $z = 1.259, p = 0.1039, N = 256$).

Second, we want to understand the mechanisms behind habitual communication. As mentioned in subsection 2.3, participants may stick with their strategy for various reasons. One mechanism suggests that participants may stick with their communication habits because they never noticed that the bias changed. The second mechanism suggests that they did notice the change, but consciously decided not to update their strategy. To see how prevalent those mechanisms are in individual decisions, we split the group of habitual participants based on whether they increase their decision time when the bias changes. Out of 112 habitual participants, 50 decrease their decision time and 62 increase it. We interpret the first group of habitual participants as evidence for habits persisting because they never noticed the change in the bias. We interpret the second group of participants as evidence for habits persisting also for participants who did notice the change in the bias.

While we cannot further break the group who did notice the change in the bias into those who stuck with their strategy to economise on the cognitive costs of reoptimising and those who did so because they believed their strategic counterpart would not change their strategy, we have some suggestive evidence for the second mechanism. More specifically, we observe that habits persist more among (64) receivers than (48) senders (Wilcoxon ranksum test, $z = 2.020, p = 0.0219, N = 256$). Given the sequential nature of the game, receivers communicating habitually can be viewed as a best response to senders communicating habitually. Thus, for some participants, habits may persist due to the belief that they interact with habitual counterparts.

Third, we provide more behavioural measures painting the profile of habitual participants.

[17] The full classification of strategies results (for both habitual and non-habitual participants, and for both part one and part two) are presented in subsection A.5. There we also perform two robustness checks of our classification procedure. First, we augment our procedure to include mixed strategies for participants who are unclassified with pure strategies. Second, we use a threshold of 80%. This threshold of 60% has been used in behavioural type analysis of sender-receiver games in (Cai and Wang, 2006; Wang et al., 2010). With the higher threshold, essentially we require an even higher automaticity. All reported patterns are qualitatively the same.

We would expect habitual participants to decide faster in the new environment. This is clearly supported by our data. When facing the new environment, habitual participants on average made decisions in 13.47 seconds whereas non-habitual participants made decisions in 16.47 seconds (Wilcoxon ranksum test, $z = 2.799, p = 0.0051, N = 256$).

We would also expect habitual participants to have lower CRT scores. CRT is a proxy for the tendency to rely on intuition versus applying deliberate thinking. Given that overriding habits requires cognitive effort, participants with higher CRT would be more likely to adapt their strategies. In line with our expectations, we find suggestive evidence that habitual participants have a CRT score of 2.06 whereas non-habitual participants have a CRT score of 2.24 (Wilcoxon ranksum test, $z = 1.729, p = 0.0838, N = 256$).

Finally, we look at the effect of the complexity of part one environment on habit formation. Aggregated, in Aligned-Frequent and Aligned-Rare 66 out of 128 participants behaved habitually compared to 46 out of 128 Conflict-Frequent and Conflict-Rare (proportion test, $z = 2.5198, p = 0.0117, N = 256$). Thus, more participants relied on habits if they started with the common-interest environment compared to conflicting-interest environment. This observation suggests that the simplicity of the common-interest environment facilitated the formation of stronger habits and is in line with psychology findings on the effect of complexity on habit formation (Wood et al., 2002; Verplanken, 2006).[18]

4 Concluding remarks

The key takeaways from this paper are: (i) habits affect strategic communication in unfamiliar environments, and (ii) reliance on communication habits is moderated by the frequency of interacting in the unfamiliar environment. Our results provide support for the conjecture that overcommunication is partially attributed to the fact that in daily interactions telling the truth and believing what you hear work well most of the time. Hence, familiarity with environments that support informative communication (outside of the lab) may lead to excessively informative communication when participants communicate in an experiment (inside the lab). By creating a counterfactual environment where communicating honestly does not pay off, we document undercommunication. We also find evidence that multiple mechanisms are at play for habitual communication, an observations that corroborates the findings of Hirmas et al. (2021) who showed that both endogenous and exogenous attention influence risk taking behaviour.

The sender-receiver cheap talk game in our experiment is arguably abstract. However, it has been extensively used in experiments on communication (Cai and Wang, 2006; Sánchez-Pagés and Vorsatz, 2007; Wang et al., 2010), and we argue it has external validity. First, lying in sender-receiver games like ours, but also in similarly abstract tasks like private die rolls, have been shown to predict behaviours such as charitable giving (Gneezy et al., 2014) and selection into public service (Hanna and Wang, 2017). It can also be manipulated with preexisting contextual cues such as criminal identity (Cohn et al., 2015) or professional affiliation (Cohn et al., 2014). Second, meta analytic evidence on a range of lying tasks, including sender-receiver

[18] This pattern is further supported by comparing decision times and time spent on feedback screens. In Aligned participants made decisions in 8.92 seconds and spent 20.02 seconds looking at the feedback screens, whereas in Conflict they made decisions in 19.31 seconds and spend 30.66 seconds on feedback screens.

games, suggest that behaviour correlates strongly across those tasks, but more importantly correlates with behaviour such as fare dodging and academic cheating (Köbis et al., 2019; Gerlach et al., 2019). Third, we have psychophysiological evidence that lying in sender-receiver games, and in similarly abstract tasks like reported spotted differences or private coin toss outcomes, correlates with pupil dilation (Wang et al., 2010), and is linked with activity in the relevant brain regions (Abe and Greene, 2014; Speer et al., 2020, 2021). Thus, we believe our results can inform our understanding of how habits affect communication outside the lab.

Communication habits imply that our everyday environment may create different predispositions. To illustrate, different occupations are characterised by different levels of preference alignment. Doctors typically have aligned preferences with their patients whereas judges often have misaligned preferences with suspects. Doctors may develop the habit of believing information whereas judges may develop the habit of mistrusting information. When communicating outside of their familiar work environment, they may carry their predisposition with them. Anecdotally, the competition for the World's Biggest Liar is held annually at the Bridge Inn in northern England. Contestants from across the world try to come up with the most convincing lie. The rules forbid lawyers and politicians from participating because "they are judged to be too skilled at telling porkies" (BBC, 2021).

Our results also suggest that habit formation may be a different process than preference formation. The effect of the environment a child grows up in on preference formation has been documented for social preferences (Cappelen et al., 2020; Kosse et al., 2020), honesty preferences (Abeler et al., 2024), and risk and time preferences (Abeler et al., 2023). (Sobel, 2013) suggested that overcommunication could be attributed to the way we learn language, which typically happens in a common-interest environment such as a family unit. While we cannot emulate developmental processes and language acquisition within the lab, we indirectly shed some light on the debate on the origins of overcommunication. Under the assumption that preferences are malleable, an arguably strong assumption for a laboratory experiment, preference formation in part one of our experiment would have made participants more or less lying averse depending on the treatment. If preference formation was the main driver, one would expect that their newly acquired preferences would lead to the same behaviour irrespective of the frequency of interacting in the new environment. We find this not to be the case. At the same time, our undercommunication finding could not be captured by theoretical models based on lying aversion (Kartik, 2009; Gneezy et al., 2018; Abeler et al., 2019) as it would require a preference *for* lying.[19]

Communication habits also have policy implications for misinformation. A wealth of evidence shows that people are not much better than chance at accurately judging the truthfulness of information (Bond Jr and DePaulo, 2006). In a recent experiment, (Serra-Garcia and Gneezy, 2021) find that conditional on judging a piece of information as truthful, senders are more likely to share it, and conditional on a piece of information being shared, receivers are more likely to

[19] A preference for lying, or equivalently a negative lying aversion, has limited power to explain large undercommunication. If we assume the utility function of a sender putting a large weight on lying, the sender may end up indirectly revealing his information by lying too much. To illustrate, if the preference for lying is sufficiently high, then when the state is very low, the sender would send the message that it is very high, and vice versa. In equilibrium, a receiver with consistent beliefs would reverse the message, so despite the extreme lying, their communication will end up being informative.

believe it. Having shown in our experiment that receivers who are mostly exposed to truthful information may form the habit of believing information, our results suggest that their habit can make receivers overly credulous and more susceptible to believing fake news and misinformation. Thus, studying the effect of habits on believing and sharing false information is an interesting avenue for future research.

More broadly, our results suggest that habit formation plays an important role in economic decision making (in our case, strategic information transmission). Thus, it is important to take into account whether a given economic situation we are studying resembles a situation with which agents may be more more familiar. Especially when we study less frequent phenomena, reliance on previously acquired habits may be a good predictor of behaviour. Myerson (1991) proposed the notion of salient perturbation, according to which behaviour in one decision environment may be predicted by decisions in a more familiar similar environment, i.e. an environment which is one salient perturbation away from the current one. Theoretical models in this direction assume that agents have access to a set of mental models, frequently referred to as analogies (Samuelson, 2001; Jehiel, 2005; Jehiel and Koessler, 2008; Jehiel, 2021). Instead of optimising for a given environment, agents locate the closest analogy to the given setting, and make a decision resembling optimal decisions for that analogy. While analogies may be acquired via different mechanisms than habit formation, such models do take into account the effect of familiar environments for decisions in unfamiliar settings.

Appendix A Additional results and robustness checks

This Appendix consists of five subsections. In subsection A.1 we list all perfect Bayesian equilibria of the game. In subsection A.2 we analyse in more detail behaviour from part one and show that findings from previous experimental literature replicate. In subsection A.3 we present econometric evidence for our main treatment effects via ordered logistic regressions. In subsection A.4 we apply the econometric method of Cai and Wang (2006) as a robustness check for our results on overcommunication and undercommunication. Finally, in subsection A.5 we present the full classification of participants in behavioural strategies from both part one and part two, and also repeat our individual behaviour analysis with a higher threshold for classifying strategies.

A.1 All perfect Bayesian equilibria of the game

Table A1 lists the complete set of all perfect Bayesian equilibria of the game for all possible values of b. The equilibria are ranked in order of informativeness as captured by the correlation between state and action.

Each row of the table represents one equilibrium. The *Messages* column describes the sender's partition of the state space. The *Actions* column describes the receiver's partition of the message space. For example, the second row is to be read as follows. The sender partitions the state space into two elements, $\{1\}$ and $\{2, 3, 4, 5\}$. If the state is 1, the sender sends the message "The state is 1". If the state is either 2, 3, 4 or 5, the senders randomly sends a message between "The state is 2", "The state is 3", "The state is 4" and "The state is 5". In

Equilibrium		Corr(S,A)	Range of values for bias b parameter				
Messages	Actions		[0, 0.22)	[0.22, 0.50)	[0.50, 0.73)	[0.73, 1.28)	$[1.28, \infty)$
{1, 2, 3, 4, 5}	{3}	0.00	✓	✓	✓	✓	✓
{1}, {2, 3, 4, 5}	{1}, {3, 4}	0.65		✓	✓	✓	
{1, 2}, {3, 4, 5}	{1, 2}, {4}	0.84	✓	✓	✓		
{1, 2, 3}, {4, 5}	{2}, {4, 5}	0.84	✓				
{1}, {2}, {3, 4, 5}	{1}, {2}, {4}	0.90	✓	✓			
{1}, {2, 3}, {4, 5}	{1}, {2, 3}, {4, 5}	0.90	✓	✓			
{1, 2}, {3}, {4, 5}	{1, 2}, {3}, {4, 5}	0.90	✓				
{1, 2}, {3, 4}, {5}	{1, 2}, {3, 4}, {5}	0.90	✓				
{1}, {2}, {3}, {4, 5}	{1}, {2}, {3}, {4, 5}	0.95	✓	✓			
{1}, {2}, {3, 4}, {5}	{1}, {2}, {3, 4}, {5}	0.95	✓				
{1}, {2, 3}, {4}, {5}	{1}, {2, 3}, {4}, {5}	0.95	✓				
{1, 2}, {3}, {4}, {5}	{1, 2}, {3}, {4}, {5}	0.95	✓				
{1}, {2}, {3}, {4}, {5}	{1}, {2}, {3}, {4}, {5}	1.00	✓	✓			

Table A1: All perfect Bayesian Nash equilibria for all positive values of b

this equilibrium, the message "The state is 1" is followed by the receiver choosing action 1. Any other message is interpreted as carrying the information that the true state is equally likely to be anywhere between 2 and 5. In that case, the best response of the receiver is to choose action 3 or action 4 with equal probabilities.

A.2 Behaviour in part one and replicating past experimental findings

This subsection serves two goals. First, it documents the successful manipulation in part one of the experiment. Second, it provides evidence replicating past findings in experiments testing the comparative statics of Crawford and Sobel (1982). For this subsection, which is based on data from part one only, we merge data from Aligned-Rare and Aligned-Conflict treatments, and from Conflict-Rare with Conflict-Frequent treatments, and refer to them as Aligned and Conflict treatments respectively.

For the successful manipulation check, we present two pieces of evidence based on correlations and decision times. First, we document that communication was more informative in the Aligned treatment compared to the Conflict treatment. Table A2 shows the correlations between states and actions, states and messages, and messages and actions in part one. The correlation between states and actions in the Aligned treatment is significantly higher than the correlation in the Conflict treatment (Wilcoxon ranksum test, $z = 4.753, p < 0.001, N = 32$). The same pattern is observed when comparing correlations between states and messages (Wilcoxon ranksum test, $z = 4.748, p < 0.001, N = 32$) and correlations between messages and actions (Wilcoxon ranksum test, $z = 4.773, p < 0.001, N = 32$).

Next we document that (i) decision times differ between treatments and decrease over rounds, and (ii) the feedback times differ between treatments and do not decrease over rounds. Table A3 shows regressions of decision times and of time spent on feedback screen on treatment and round. The regressions are repeated once on individual and once on matching group level.

We now compare our results to the previous literature. We observe overcommunication in the Conflict treatment as all correlations are significantly positive (Wilcoxon signrank test, $z = 3.516$, $p < 0.001$, $N = 16$).[20] Rows 3-6 of Table A2 show the correlations obtained from two

[20] Technically, correlations in Aligned are all significantly lower than 1.00, but this is driven by a ceiling effect.

Paper	Treatment	Corr(S,A)	Corr(S,M)	Corr(M,A)	Predicted
Current	Aligned	0.953	0.967	0.982	1.000
	Conflict	0.387	0.528	0.647	0.000
Cai and Wang (2006)	Aligned	0.876	0.916	0.965	1.000
	Conflict	0.207	0.391	0.542	0.000
Wang et al. (2010)	Aligned	0.860	0.930	0.920	1.000
	Conflict	0.320	0.340	0.580	0.000

Table A2: Correlations between states, messages, and actions

	Decision Time		Feedback Time	
	Group	Individual	Group	Individual
Aligned	-11.31***	-10.87***	-11.06***	-11.33***
	(1.49)	(1.22)	(2.11)	(1.16)
Round	-0.41***	-0.38***	0.04	0.04
	(0.05)	(0.03)	(0.06)	(0.06)
Controls	No	Yes	No	Yes
Observations	960	7680	960	7680

Controls: Age, Gender, Study, Risk, CRT, Trust.

Standard errors clustered on matching group level (32 clusters).

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table A3: Decision and feedback times in part one

previous papers that used the same configuration as the current paper. Similar to us, they also observe overcommunication even though only a babbling equilibrium exists. Thus, behaviour from part one in the Conflict treatment supports past findings of overcommunication.

A.3 Econometric tests for treatment effects

In this subsection, we are interested in testing whether starting from the Aligned environment in part one leads to more informative communication when interacting in the new environment in part two compared to starting from the Conflict environment. To do so, we estimate ordered logistic regressions of receiver action on state and interact state with part one environment. A significant interaction (*State*×*Aligned*) translates to more informative communication after Aligned compared to after Conflict. We estimate separate regressions for when the new environment occurs rarely or frequently, and separate for early and late rounds that it does so. Each regression is estimated using individual choices with errors clustered at the matching group level. The regressions control for risk, CRT, trust towards strangers, and demographics.

Our results reveal a treatment effect when the new environment occurs rarely (columns 1 and 2) and a null effect when the new environment occurs frequently (columns 3 and 4). Thus Result 1 and Result 2 from the main text are robust.

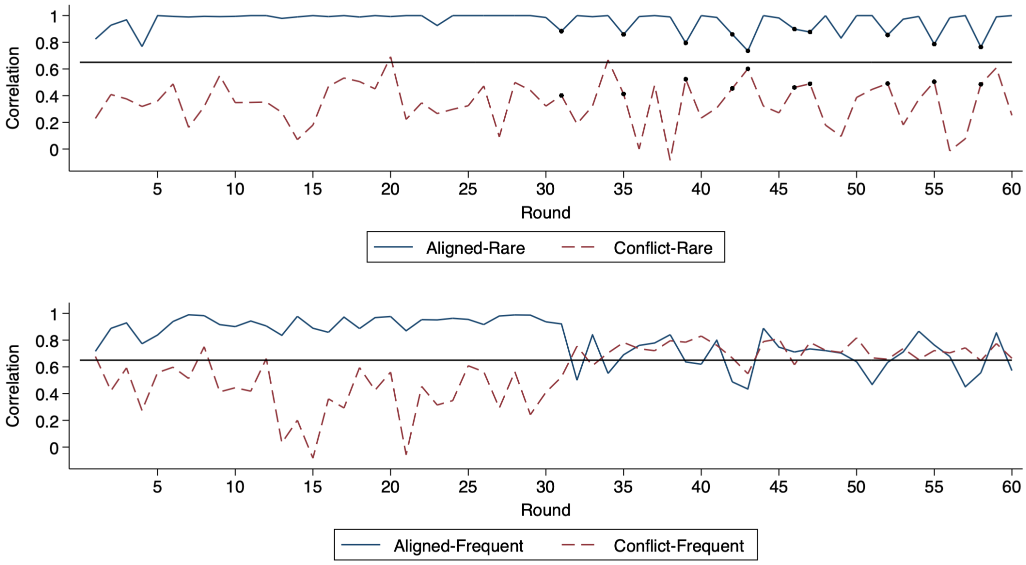
We visualise the results in Figure A1 where we plot the time series of correlations across all rounds of the experiment. In the upper panel we see that in the Rare treatment, correlations after Aligned always exceed correlations after Conflict. In the lower panel, we see that this is

	Receiver's action			
	Rare Early	Rare Late	Frequent Early	Frequent Late
State	1.16***	1.16***	1.49***	1.64***
	(0.20)	(0.23)	(0.22)	(0.16)
State×Aligned	0.36***	0.33**	0.09	-0.05
	(0.09)	(0.11)	(0.11)	(0.09)
Round	0.02	-0.02	-0.05	0.08
	(0.02)	(0.03)	(0.08)	(0.07)
Controls	Yes	Yes	Yes	Yes
Observations	640	640	640	640

Controls: Age, Gender, Study, Risk, CRT, Trust.
Standard errors clustered on matching group level (16 clusters).
Significance levels: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table A4: Ordered logistic regression of action on state

not the case in the Frequent treatments, as correlations in part two overlap. By focusing only on the first 30 rounds of the graphs, the reader can also observe the successful manipulation discussed extensively in subsection A.2.



In the upper figure for Aligned-Rare and Conflict-Rare, black dots indicate the rounds when b=1.

Figure A1: Correlations over rounds

A.4 Econometric tests for overcommunication and undercommunication

This subsection provides a robustness check for Result 3 on overcommunication and undercommunication presented in subsection 3.2. To do so, we use the regression method developed by Cai and Wang (2006, Result 3).

The method uses a standard regression as a starting point. Consider a model $Y = \alpha + \beta X + \epsilon$. The estimator for β is given by $b = \frac{SD_Y}{SD_X} \times Corr(X, Y)$, where SD_Y, SD_X are the sample standard deviations of X and Y respectively, and $Corr(X, Y)$ is the correlation between X and Y. To test whether the estimated correlation differs from a theoretical one (denoted by σ_{XY}), it suffices to estimate the adjusted model $Y - r_{XY}X = \alpha + \beta X + \epsilon$, where $r_{XY} = \frac{SD_Y}{SD_X} \times \sigma_{XY}$. The t-test on the estimate of β in the adjusted model allows us to precisely test whether $Corr(X, Y) = \sigma_{XY}$. We estimate those regressions separately for each of the four treatments, and separately for early and late rounds. For all regressions, we use the correlation of the most informative perfect Bayesian equilibrium as the theoretical prediction ($\sigma_{XY} = 0.650$).

	Action							
	CR Early	CR Late	AR Early	AR Late	CF Early	CF Late	AF Early	AF Late
State	-0.14*	-0.15**	0.19***	0.17***	0.03	0.09**	0.10	0.07
	(0.05)	(0.05)	(0.03)	(0.03)	(0.04)	(0.03)	(0.05)	(0.04)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	320	320	320	320	320	320	320	320

Controls: Age, Gender, Study, Risk, CRT, Trust.

Standard errors clustered on participant level (64 clusters).

Significance levels: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table A5: Regressions of (adjusted) action on state

When the new environment occurs rarely (first four columns), we see significant differences from equilibrium predictions. When participants started in Conflict (columns 1 and 2), we observe undercommunication as the coefficient on state is negative. When participants started in Aligned (columns 3 and 4), we observe overcommunication as the coefficient on state is positive. We observe no significant differences when the new environment occurs frequently (with the exception of column 6). Thus, Result 3 is robust.

A.5 Full classification of individual behavioural strategies

This subsection serves two goals. First, it provides more details on the classification procedure of behavioural strategies. Second, it briefly shows that the results discussed in subsection 3.3 are robust to a stricter classification procedure.

As discussed in the main text, we consider all possible strategies for senders and receivers. This means $5^3 = 3,125$ strategies for senders and equally as many for receivers. We do this process twice; once for rounds 11-20 from part one, and once for the 10 rounds from part two where $b = 1$. We assign to each participant in each part the strategy that matches the majority of their decisions as long as it matches at least 60% of their decisions. This procedure classifies into behavioural strategies 235/256 participants in part one and 242/256 in part two. Table A6 lists all the pure strategies that have at least one participant classified in either part one or part two. The table also includes the short name of each strategy which will be used later in this subsection.

Some of the unclassified participants could have formed the habit of being unpredictable by using a mixed strategy. To account for the possibility of habitual mixing, we augment our procedure with an additional step which attempts to correct for this limitation. We estimate a

Strategy	Coding
Tell the truth	Truth
Exaggerate state by 1	s+1
Exaggerate state by 2	s+2
Exaggerate state by 3	s+3
Always send message 4	m=4
Always send message 5	m=5

(a) Sender strategies

Strategy	Coding
Follow message	Believe
Discount message by 1	m-1
Discount message by 2	m-2
One more than message	m+1
Always choose action 3	a=3
Always choose action 4	a=4

(b) Receiver strategies

Table A6: All pure strategies used by participants

regression of choice on cue including data from both parts and incorporate an interaction effect to allow for different slopes across parts. For senders, the cue is the state and the choice is the message. For receivers, the cue is the message and the choice is the action. Formally, we estimate the following regression:

$$\text{Choice}_i = \beta_0 + \beta_1 * \text{Cue}_i + \beta_2 * \text{Part}_i + \beta_3 * \text{Part}_i * \text{Cue}_i + \epsilon$$

If β_2 and β_3 are jointly significant, then the participant changed strategy. If not, then the participant used the same strategy and is classified as habitual. Our second step essentially equates habitual communication with making (statistically) similarly informative choices between part one and part two.

In total 106 participants are classified as habitual with pure strategies, and 6 participants are classified as habitual with mixed strategies, resulting in 112 habitual participants in total. Table A7 and Table A8 present the full classification of behavioural strategies for senders and receivers respectively. Habitual participants are on the diagonal of the tables.

As a first robustness check, we briefly comment on the fact that some combinations of strategies could be considered habitual in the level-k sense. To illustrate, a level-1 sender will naively assume that the receiver will follow their message, and exaggerate the state by 2 when $b = 2$ and by 1 when $b = 1$. Similarly, a level-2 receiver would assume senders behave as level-1, and consequently choose an action two lower than the message when $b = 2$ and one lower than the message when $b = 1$. With this in mind, 8 senders and 7 receivers in Conflict-Rare, and 10 senders and 7 receivers in Conflict-Frequent could also be classified as habitual. Qualitatively our results are similar if we take level-k classifications into account.[21]

As a second robustness check, we increase the threshold for a strategy to be assigned to a participant if the strategy matches at least 80% of their decisions. By raising the threshold from 60% to 80%, we essentially require higher automaticity. Out of 112 habitual participants, 102 are still classified as habitual with this higher threshold. This suggests that our procedure is relatively robust to the chosen threshold.

[21] If we only allow classifications based on level-k behaviour, we find that the distribution of level-k types differs between aligned and conflict environments ($\chi^2 = 10.789$, p-value=0.001, $N = 256$). This observation corroborates literature suggesting that strategic reasoning, as measured by level-k, is not a fixed characteristic of an agent, but it changes across situations or over time (Agranov et al., 2012; Georganas et al., 2015; Alaoui et al., 2020).

	Truth		s+1		s+2		s+3		m=4		m=5		Mix		None	
Truth	11	12	17	17	1	1									3	
	2	6													1	
s+1				1												
	1	2	1	4												
s+2																
	1		8	10	4						1					
s+3																
					1	1										
m=4																
	1			1												
m=5																
	1					2	1				2					
Mix														1		
													4			
None																
		1	1	4		1			1		1					

Rows represent strategies used in part one. Columns represent strategies used in part two.
Each cell is split into four mini-cells with different shading.
In each cell: top-left is A-R, top-right is A-F, bottom-left is C-R, bottom-right is C-F.
A-R=Aligned-Rare, A-F=Aligned-Frequent, C-R=Conflict-Rare, C-F=Conflict-Frequent.

Table A7: Classification of sender strategies

To illustrate the consistency of the observations from subsection 3.3, when increasing the threshold we find: i) more habitual participants when the new environment is rare compared to frequent (57 vs 45, proportion test, $z = 1.53, p = 0.0628, N = 256$), (ii) 40 participants who decreased their decision time and 62 that increased their decision time, (iii) more habitual receivers than senders (62 vs 40, proportion test, $z = 2.81, p = 0.0025, N = 256$), (iv) habitual participants making faster decisions compared to non-habitual (12.43 seconds vs 16.97 seconds, Wilcoxon ranksum test, $z = 4.204, p < 0.0001, N = 256$), (v) habitual participants having slightly lower CRT scores (2.06 vs 2.23, Wilcoxon ranksum test, $z = 1.518, p = 0.1291, N = 256$), and (vi) more habitual participants after aligned environment compared to conflicting (57 vs 45, proportion test, $z = 1.53, p = 0.0625, N = 256$).

Appendix B Instructions

Welcome to the session

Welcome!

Thank you for participating in this study. Please make sure that you are in the Zoom meeting throughout the experiment. You were admitted to the session from the waiting room, renamed, and send back to the waiting room. This was to ensure your privacy. If you have any ques-

	Believe		m-1		m-2		m+1		a=3		a=4		Mix		None	
Believe	25	16	4	10		2									3	3
	6	6		3												1
m-1																
		2	4	1	1										1	
m-2																
	3	1	7	7	2	1									1	
m+1																
	1															
a=3																
			1							2						
a=4																
															1	
Mix																
														1		
None		1														
		4	1	3			3									

Rows represent strategies used in part one. Columns represent strategies used in part two.
Each cell is split into four mini-cells with different shading.
In each cell: top-left is A-R, top-right is A-F, bottom-left is C-R, bottom-right is C-F.
A-R=Aligned-Rare, A-F=Aligned-Frequent, C-R=Conflict-Rare, C-F=Conflict-Frequent.

Table A8: Classification of receiver strategies

tions, you can message the experiment during the experiment. The Zoom session only allows participants to message the experimenter. Any question you ask and the answer from the experimenter will **not** be shown to any other participant. Please keep your video off and stay muted throughout the experiment.

Payment registration

Please enter your IBAN below. This will be used for payment after the experiment. You will **not** be able to change this at a later point. We will delete this number after making the payment.

B.1 Overview of the experiment

Welcome!

Welcome to this experiment. Please read the following instructions carefully. We ask that you do not communicate with other participants during the experiment. The use of mobile phones is not allowed during this experiment. If you have any questions, or need assistance of any kind, at any time, please message the experimenter privately in the Zoom session and he/she will assist you. The data collected throughout this experiment does not include your name or any other information that would allow your identification. All of the data you provide during the experiment cannot be traced back to you.

Your earnings in today's session will be paid to you at the end of the experiment. Your earnings will depend on your own and other participants' decisions. You will play **60** rounds in total. For each round, your earnings will be in points. At the end of the experiment, your **accumulated** points will be converted to euros at a rate of 1 euro per 200 points. You will receive your earnings at the end of the experiment at the bank account you provided.

In the next page you will receive the relevant instructions. Thank you for your participation.

B.2 Rules of the sender-receiver game

Please read the following instructions carefully

Matching & roles

In each round, all participants are matched in pairs. One participant within a pair has the role of player A and the other participant has the role of player B. The matching scheme is chosen to guarantee the following:

- In each round you will be randomly matched to another participant.
- You will never learn with whom you are matched with.
- You will never be paired to the same participant in subsequent rounds.
- You will always have the same role in all rounds.

Sequence of actions

1. In each round of the experiment, the computer will randomly roll a die with numbers between 1 and 5. All numbers are equally likely. This outcome of the die is called the *state*. Player A will observe the state, whereas player B will not.

2. Player A moves first and has to choose between the following 5 options.

   ○ Send the message "The state is 1"
   ○ Send the message "The state is 2"
   ○ Send the message "The state is 3"
   ○ Send the message "The state is 4"
   ○ Send the message "The state is 5"

   If player A decides to send a message, it does not have to match the state. This is the only decision of player A.

3. Player B will observe the message and choose an action between 1 and 5. The decision of player B ends the round.

Earnings

In each round you can earn or lose points. The earnings of both players **depend only** on the state and the action of player B. The earnings **do not depend** on the message sent by player A. The earnings of both players for all possible combinations of state and action will be provided to you in a table. The table will be shown to both of you in the decision screen.

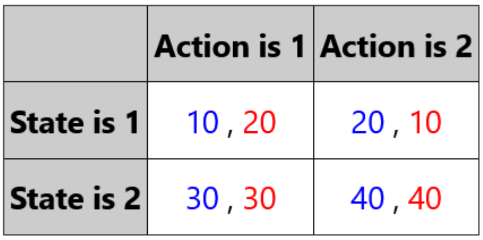
B.3 Understanding questions

Each cell of the table contains two numbers which correspond to the earnings of the two players.

- For player A, the earnings are the number on the left (shown in blue).
- For player B, the earnings are the number on the right (shown in red).

Remember that earnings depend **only** on the combination of state and action and **not** on the message.

Below there is an example of such a table to make you familiar with the format. All the scenarios described in the questions are purely hypothetical. Answering all questions correctly will ensure you fully understand the rules of the game and how points are earned.



	Action is 1	**Action is 2**
State is 1	10 , 20	20 , 10
State is 2	30 , 30	40 , 40

1. The state is 1. Player A send the message “The state is 1”. Player B chose action 2. **What are the earnings of each player?**

   ○ Player A gets 10 and player B gets 20

   ○ Player A gets 20 and player B gets 10

   ○ Player A gets 30 and player B gets 30

   ○ Player A gets 40 and player B gets 40

2. The state is 1. Player A send the message “The state is 2”. Player B chose action 2. **What are the earnings of each player?**

   ○ Player A gets 10 and player B gets 20

   ○ Player A gets 20 and player B gets 10

   ○ Player A gets 30 and player B gets 30

   ○ Player A gets 40 and player B gets 40

3. The state is 2. Player A send the message “The state is 2”. Player B chose action 2. **What are the earnings of each player?**

○ Player A gets 10 and player B gets 20

○ Player A gets 20 and player B gets 10

○ Player A gets 30 and player B gets 30

○ Player A gets 40 and player B gets 40

4. The state is 2. Player A send the message “The state is 1”. Player B chose action 2. **What are the earnings of each player?**

   ○ Player A gets 10 and player B gets 20

   ○ Player A gets 20 and player B gets 10

   ○ Player A gets 30 and player B gets 30

   ○ Player A gets 40 and player B gets 40

5. When player A chooses the message to send to player B, both players know the state. **Is this statement True of False?**

   ○ True

   ○ False

6. Player A can send the message “The state is 2” when the state is 1. **Is this statement True of False?**

   ○ True

   ○ False

7. Player A sent the message “I don’t want to send a message” when the state is 1. Player B chose action 2. **What are the earnings of each player?**

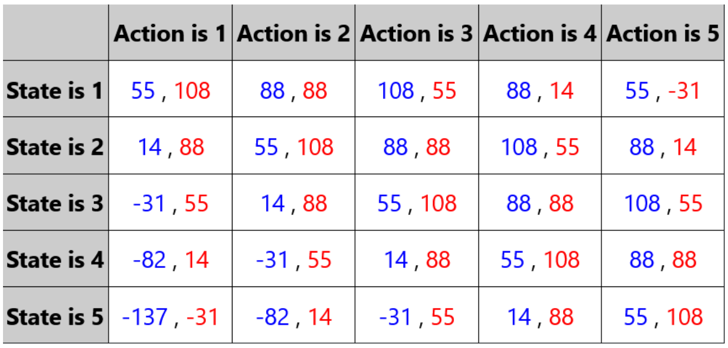
Click the “Check” button below to check your answers. You can only proceed to the next page if all answers are correct.

B.4 Decision screen

Round X of 60

Below you see the table containing the earnings for both players for every combination of state and action.

- For player A, the earnings are the number on the left (shown in blue).
- For player B, the earnings are the number on the right (shown in red).



	Action is 1	Action is 2	Action is 3	Action is 4	Action is 5
State is 1	55 , 108	88 , 88	108 , 55	88 , 14	55 , -31
State is 2	14 , 88	55 , 108	88 , 88	108 , 55	88 , 14
State is 3	-31 , 55	14 , 88	55 , 108	88 , 88	108 , 55
State is 4	-82 , 14	-31 , 55	14 , 88	55 , 108	88 , 88
State is 5	-137 , -31	-82 , 14	-31 , 55	14 , 88	55 , 108

[SENDER] You are **player A**. The randomly drawn state is DIE PHOTO (**2**). Please choose a message to send to player B by clicking the corresponding button below.

[RECEIVER AFTER ACTIVE SENDER] You are **player B**. Player A sent you the message "The state is 5". Please choose a message to send to player B by clicking the corresponding button below.

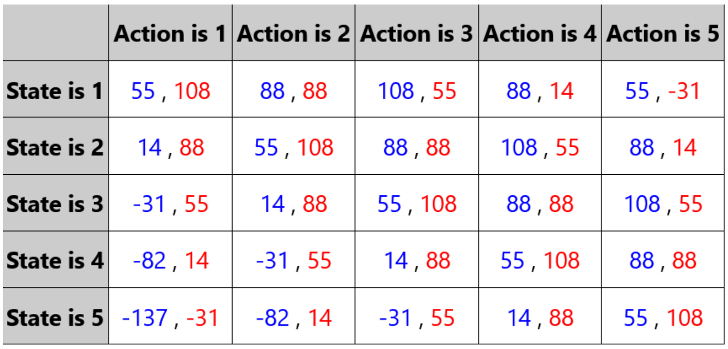
[RECEIVER AFTER INACTIVE SENDER] You are **player B**. Player A was inactive in this round due to technical/connectivity issues. Hence, click Next to proceed.

B.5 Feedback screen

Results from round X of 60

Below you see the table containing the earnings for both players for every combination of state and action.

- For player A, the earnings are the number on the left (shown in blue).
- For player B, the earnings are the number on the right (shown in red).



	Action is 1	Action is 2	Action is 3	Action is 4	Action is 5
State is 1	55 , 108	88 , 88	108 , 55	88 , 14	55 , -31
State is 2	14 , 88	55 , 108	88 , 88	108 , 55	88 , 14
State is 3	-31 , 55	14 , 88	55 , 108	88 , 88	108 , 55
State is 4	-82 , 14	-31 , 55	14 , 88	55 , 108	88 , 88
State is 5	-137 , -31	-82 , 14	-31 , 55	14 , 88	55 , 108

[PLAYER ACTIVE, PARTNER ACTIVE] The state was 2. Player A send the message "The state is 5". Player B chose action 3.

[SENDER ONLY] You were **player A**. Therefore, in this round you earned 88 points.

[RECEIVER ONLY] You were **player B**. Therefore, in this round you earned 88 points.

[PLAYER ACTIVE, PARTNER INACTIVE] Your partner was inactive in this round so you automatically earned 100 points.

[PLAYER INACTIVE] You were inactive in this round and automatically earned 0 points.

B.6 Survey

Lottery Task

In the following task, **5 different lotteries** will be presented on your screen. In each of these lotteries, **both rewards** A and B are **equally likely**, i.e. have a probability of exactly 50%. The rewards are denoted in points.

You are asked to **choose exactly one** of the lotteries, which subsequently will be implemented. A random generator will determined whether you win reward A or reward B, respectively. At the end of the experiment, your reward will be added to your earnings.

	Reward A	Reward B	
No.	50% Probability	50% Probability	Your Choice
1.	140	140	
2.	120	180	
3.	100	220	
4.	80	260	
5.	60	300	
6.	10	350	

CRT elicitation

Please answer the following questions. Each correct answer is worth 50 points.

1. The ages of Mark and Adam add up to 28 years in total. Mark is 20 years older than Adam. How many years old is Adam?

2. If it takes 10 seconds for 10 printers to print out 10 pages of paper, how many seconds will it take for 50 printers to print out 50 pages of paper?

3. On a loaf of bread, there is a patch of mould. Every day the patch doubles in size. If it takes 12 days for the patch to cover the entire load of bread, how many days would it take for the patch to cover half the loaf of bread?

Trust attitudes

Please answer the following questions.

- When I communicate with strangers, I tell them the truth.
  (Strongly disagree, Disagree, Neither agree or disagree, Agree, Strongly agree)

- When I communicate with strangers, they tell me the truth.
  (Strongly disagree, Disagree, Neither agree or disagree, Agree, Strongly agree)

Demographics

Please answer the following questions.

- Please indicate your age.

- Please indicate your field of study.
  (Economics, Social Sciences, Natural Sciences, Humanities, Applied Sciences, Other)
- Please indicate your gender.
  (Male, Female, Prefer not to answer)

B.7 Payment information and debriefing

Thank you!

The experiment is completed. Thank you for your participation.

From the main game, you earned in total 94 points. For the other tasks you additionally earned 154 points. The exchange rate is €1 for 200 points, so you earned €0.77.

You will receive your payment to your bank account using the IBAN you provided in the beginning of the experiment. You can now leave the Zoom session and close your browser.

References

Abe, N. and Greene, J. D. (2014). Response to anticipated reward in the nucleus accumbens predicts behavior in an independent test of honesty. *Journal of Neuroscience*, 34(32):10564–10572.

Abeler, J., Falk, A., and Kosse, F. (2024). Malleability of preferences for honesty. *The Economic Journal*.

Abeler, J., Fosgaard, T. R., and Gårn Hansen, L. (2023). The effect of the social environment during childhood on preferences in adulthood. Technical report, IFRO Working Paper.

Abeler, J., Nosenzo, D., and Raymond, C. (2019). Preferences for truth-telling. *Econometrica*, 87(4):1115–1153.

Acland, D. and Levy, M. R. (2015). Naiveté, projection bias, and habit formation in gym attendance. *Management Science*, 61(1):146–160.

Agranov, M., Potamites, E., Schotter, A., and Tergiman, C. (2012). Beliefs and endogenous cognitive levels: An experimental study. *Games and Economic Behavior*, 75(2):449–463.

Alaoui, L., Janezic, K. A., and Penta, A. (2020). Reasoning about others' reasoning. *Journal of Economic Theory*, 189(1):1–51.

Amaya, K. A. and Smith, K. S. (2018). Neurobiology of habit formation. *Current opinion in behavioral sciences*, 20:145–152.

Angelova, V. and Regner, T. (2013). Do voluntary payments to advisors improve the quality of financial advice? An experimental deception game. *Journal of economic behavior & organization*, 93:205–218.

Arbel, Y., Bar-El, R., Siniver, E., and Tobol, Y. (2014). Roll a die and tell a lie–what affects honesty? *Journal of Economic Behavior & Organization*, 107(1):153–172.

BBC (2021). World's biggest liar championship. Source: http://www.bbc.com/storyworks/a-year-of-great-events/worlds-biggest-liar-championship. Accessed: 2021-06-03.

Belot, M. and van de Ven, J. (2019). Is dishonesty persistent? *Journal of Behavioral and Experimental Economics*, 83:1–9.

Blume, A., Lai, E. K., and Lim, W. (2020). Strategic information transmission: A survey of experiments and theoretical foundations. In *Handbook of Experimental Game Theory*. Edward Elgar Publishing.

Bond Jr, C. F. and DePaulo, B. M. (2006). Accuracy of deception judgments. *Personality and social psychology Review*, 10(3):214–234.

Bordalo, P., Gennaioli, N., and Shleifer, A. (2022). Salience. *Annual Review of Economics*, 14(1):521–544.

Buyalskaya, A., Ho, H., Milkman, K. L., Li, X., Duckworth, A. L., and Camerer, C. (2023). What can machine learning teach us about habit formation? Evidence from exercise and hygiene. *Proceedings of the National Academy of Sciences*, 120(17).

Byrne, D. P., Goette, L., Martin, L. A., Miles, A., Jones, A., Schob, S., Staake, T., and Tiefenbeck, V. (2023). How nudges create habits: Theory and evidence from a field experiment. Technical report, University of Bonn and University of Mannheim, Germany.

Cai, H. and Wang, J. T.-Y. (2006). Overcommunication in strategic information transmission games. *Games and Economic Behavior*, 56(1):7–36.

Camerer, C., Xin, Y., and Zhao, C. (2024). A neural autopilot theory of habit: Evidence from consumer purchases and social media use. *Journal of the Experimental Analysis of Behavior*, 121(1):108–122.

Cappelen, A., List, J., Samek, A., and Tungodden, B. (2020). The effect of early-childhood education on social preferences. *Journal of Political Economy*, 128(7):2739–2758.

Cassar, A., d'Adda, G., and Grosjean, P. (2014). Institutional quality, culture, and norms of cooperation: Evidence from behavioral field experiments. *The Journal of Law and Economics*, 57(3):821–863.

Chakraborty, A. and Harbaugh, R. (2014). Persuasive puffery. *Marketing Science*, 33(3):382–400.

Charness, G. and Gneezy, U. (2009). Incentives to exercise. *Econometrica*, 77(3):909–931.

Chen, D. L., Schonger, M., and Wickens, C. (2016). oTree–An open-source platform for laboratory, online, and field experiments. *Journal of Behavioral and Experimental Finance*, 9(1):88–97.

Chen, Y., Kartik, N., and Sobel, J. (2008). Selecting cheap-talk equilibria. *Econometrica*, 76(1):117–136.

Cohn, A., Fehr, E., and Maréchal, M. A. (2014). Business culture and dishonesty in the banking industry. *Nature*, 516(7529):86–89.

Cohn, A., Maréchal, M. A., and Noll, T. (2015). Bad boys: How criminal identity salience affects rule violation. *The Review of Economic Studies*, 82(4):1289–1308.

Condorelli, D. and Furlan, M. (2023). Cheap talking algorithms. Working paper, University of Warwick.

Coppock, A. and Green, D. P. (2016). Is voting habit forming? new evidence from experiments and regression discontinuities. *American Journal of Political Science*, 60(4):1044–1062.

Costa-Gomes, M. A. and Crawford, V. P. (2006). Cognition and behavior in two-person guessing games: An experimental study. *American economic review*, 96(5):1737–1768.

Crawford, V. P. and Sobel, J. (1982). Strategic information transmission. *Econometrica*, 50(6):1431–1451.

Dal Bó, P. and Fréchette, G. R. (2011). The evolution of cooperation in infinitely repeated games: Experimental evidence. *American Economic Review*, 101(1):411–29.

de Groot Ruiz, A., Offerman, T., and Onderstal, S. (2015). Equilibrium selection in experimental cheap talk games. *Games and Economic Behavior*, 91(1):14–25.

de Haan, T., Offerman, T., and Sloof, R. (2015). Money talks? An experimental investigation of cheap talk and burned money. *International Economic Review*, 56(4):1385–1426.

de Mel, S., McIntosh, C., and Woodruff, C. (2013). Deposit collecting: Unbundling the role of frequency, salience, and habit formation in generating savings. *American Economic Review*, 103(3):387–92.

Dickhaut, J. W., McCabe, K. A., and Mukherji, A. (1995). An experimental study of strategic information transmission. *Economic Theory*, 6(3):389–403.

Duffy, J. and Fehr, D. (2018). Equilibrium selection in similar repeated games: Experimental evidence on the role of precedents. *Experimental Economics*, 21:573–600.

Eckel, C. C. and Grossman, P. J. (2002). Sex differences and statistical stereotyping in attitudes toward financial risk. *Evolution and human behavior*, 23(4):281–295.

Engl, F., Riedl, A., and Weber, R. (2021). Spillover effects of institutions on cooperative behavior, preferences, and beliefs. *American Economic Journal: Microeconomics*, 13(4):261–299.

Enke, B., Graeber, T., Oprea, R., and Yang, J. (2024). Behavioral attenuation. Working paper, Harvard University and University of California, Santa Barbara.

Falk, A., Fischbacher, U., and Gächter, S. (2013). Living in two neighborhoods—social interaction effects in the laboratory. *Economic Inquiry*, 51(1):563–578.

Farrell, J. (1993). Meaning and credibility in cheap-talk games. *Games and Economic Behavior*, 5(4):514–531.

Frederick, S. (2005). Cognitive reflection and decision making. *Journal of Economic perspectives*, 19(4):25–42.

Fujiwara, T., Meng, K., and Vogl, T. (2016). Habit formation in voting: Evidence from rainy elections. *American Economic Journal: Applied Economics*, 8(4):160–88.

Gabaix, X. (2019). Behavioral inattention. In *Handbook of behavioral economics: Applications and foundations 1*, volume 2, pages 261–343. Elsevier.

Gardete, P. M. (2013). Cheap-talk advertising and misrepresentation in vertically differentiated markets. *Marketing Science*, 32(4):609–621.

Georganas, S., Healy, P. J., and Weber, R. A. (2015). On the persistence of strategic sophistication. *Journal of Economic Theory*, 159(1):369–400.

Gerber, A., Green, D., and Shachar, R. (2003). Voting may be habit-forming: evidence from a randomized field experiment. *American Journal of Political Science*, 47(3):540–550.

Gerlach, P., Teodorescu, K., and Hertwig, R. (2019). The truth about lies: A meta-analysis on dishonest behavior. *Psychological bulletin*, 145(1):1.

Glaeser, E. L., Laibson, D. I., Scheinkman, J. A., and Soutter, C. L. (2000). Measuring trust. *The quarterly journal of economics*, 115(3):811–846.

Glimcher, P. W. (2022). Efficiently irrational: deciphering the riddle of human choice. *Trends in cognitive sciences*, 26(8):669–687.

Gneezy, U., Imas, A., and Madarász, K. (2014). Conscience accounting: Emotion dynamics and social behavior. *Management Science*, 60(11):2645–2658.

Gneezy, U., Kajackaite, A., and Sobel, J. (2018). Lying aversion and the size of the lie. *American Economic Review*, 108(2):419–53.

Gravert, C. and Collentine, L. O. (2021). When nudges aren't enough: Norms, incentives and habit formation in public transport usage. *Journal of Economic Behavior & Organization*, 190:1–14.

Guo, T., Sriram, S., and Manchanda, P. (2020). "Let the sunshine in": The impact of industry payment disclosure on physician prescription behavior. *Marketing Science*, 39(3):516–539.

Hanna, R. and Wang, S.-Y. (2017). Dishonesty and selection into public service: Evidence from india. *American Economic Journal: Economic Policy*, 9(3):262–290.

Havranek, T., Rusnak, M., and Sokolova, A. (2017). Habit formation in consumption: A meta-analysis. *European Economic Review*, 95(1):142–167.

Hirmas, A., Engelmann, J., and van der Weele, J. J. (2021). Individual and contextual effects of attention in risky choice. Discussion paper 2021-031/I, Tinbergen Institute.

Holm, H. J. and Kawagoe, T. (2010). Face-to-face lying–An experimental study in sweden and japan. *Journal of Economic Psychology*, 31(3):310–321.

Hugh-Jones, D. (2016). Honesty, beliefs about honesty, and economic growth in 15 countries. *Journal of Economic Behavior & Organization*, 127(1):99–114.

Hurkens, S. and Kartik, N. (2009). Would I lie to you? on social preferences and lying aversion. *Experimental Economics*, 12(2):180–192.

Hussam, R., Rabbani, A., Reggiani, G., and Rigol, N. (2022). Rational habit formation: experimental evidence from handwashing in india. *American Economic Journal: Applied Economics*, 14(1):1–41.

Inderst, R. and Ottaviani, M. (2012). Financial advice. *Journal of Economic Literature*, 50(2):494–512.

Innes, R. and Arnab, M. (2013). Is dishonesty contagious? *Economic Inquiry*, 51(1):722–734.

Ioannidis, K. (2020). Overcommunication: The role of past experience. Preregistration, American Economic Association's registry for randomized controlled trials. `https://www.socialscienceregistry.org/trials/6387`.

Ito, K., Ida, T., and Tanaka, M. (2018). Moral suasion and economic incentives: Field experimental evidence from energy demand. *American Economic Journal: Economic Policy*, 10(1):240–267.

Jehiel, P. (2005). Analogy-based expectation equilibrium. *Journal of Economic theory*, 123(2):81–104.

Jehiel, P. (2021). Communication with forgetful liars. *Theoretical Economics*, 16(2):605–638.

Jehiel, P. and Koessler, F. (2008). Revisiting games of incomplete information with analogy-based expectations. *Games and Economic Behavior*, 62(2):533–557.

Kartik, N. (2009). Strategic communication with lying costs. *The Review of Economic Studies*, 76(4):1359–1395.

Kawagoe, T. and Takizawa, H. (2009). Equilibrium refinement vs. level-k analysis: An experimental study of cheap-talk games with private information. *Games and Economic Behavior*, 66(1):238–255.

Knez, M. and Camerer, C. (2000). Increasing cooperation in prisoner's dilemmas by establishing a precedent of efficiency in coordination games. *Organizational behavior and human decision processes*, 82(2):194–216.

Köbis, N. C., Verschuere, B., Bereby-Meyer, Y., Rand, D., and Shalvi, S. (2019). Intuitive honesty versus dishonesty: Meta-analytic evidence. *Perspectives on Psychological Science*, 14(5):778–796.

Kosse, F., Deckers, T., Pinger, P., Schildberg-Hörisch, H., and Falk, A. (2020). The formation of prosociality: causal evidence on the role of social environment. *Journal of Political Economy*, 128(2):434–467.

Lafky, J., Lai, E. K., and Lim, W. (2022). Preferences vs. strategic thinking: An investigation of the causes of overcommunication. *Games and Economic Behavior*, 136:92–116.

Lally, P., Van Jaarsveld, C. H., Potts, H. W., and Wardle, J. (2010). How are habits formed: Modelling habit formation in the real world. *European journal of social psychology*, 40(6):998–1009.

Li, X., Özer, Ö., and Subramanian, U. (2022). Are we strategically naïve or guided by trust and trustworthiness in cheap-talk communication? *Management Science*, 68(1):376–398.

Maćkowiak, B., Matějka, F., and Wiederholt, M. (2023). Rational inattention: A review. *Journal of Economic Literature*, 61(1):226–273.

Matthews, S. A., Okuno-Fujiwara, M., and Postlewaite, A. (1991). Refining cheap-talk equilibria. *Journal of Economic Theory*, 55(2):247–273.

McCarter, M. W., Samek, A., and Sheremeta, R. M. (2014). Divided loyalists or conditional cooperators? Creating consensus about cooperation in multiple simultaneous social dilemmas. *Group & Organization Management*, 39(6):744–771.

Meredith, M. et al. (2009). Persistence in political participation. *Quarterly Journal of Political Science*, 4(3):187–209.

Myerson, R. B. (1991). *Game Theory: Analysis of Conflict.* Harvard University Press.

Pascual-Ezama, D., Fosgaard, T. R., Cardenas, J. C., Kujal, P., Veszteg, R., de Liaño, B. G.-G., Gunia, B., Weichselbaumer, D., Hilken, K., Antinyan, A., et al. (2015). Context-dependent cheating: Experimental evidence from 16 countries. *Journal of Economic Behavior & Organization*, 116(1):379–386.

Payzan-LeNestour, E. and Woodford, M. (2022). Outlier blindness: A neurobiological foundation for neglect of financial risk. *Journal of Financial Economics*, 143(3):1316–1343.

Peysakhovich, A. and Rand, D. G. (2016). Habits of virtue: Creating norms of cooperation and defection in the laboratory. *Management Science*, 62(3):631–647.

Royer, H., Stehr, M., and Sydnor, J. (2015). Incentives, commitments, and habit formation in exercise: Evidence from a field experiment with workers at a fortune-500 company. *American Economic Journal: Applied Economics*, 7(3):51–84.

Samuelson, L. (2001). Analogies, adaptation, and anomalies. *Journal of Economic Theory*, 97(2):320–366.

Sánchez-Pagés, S. and Vorsatz, M. (2007). An experimental study of truth-telling in a sender–receiver game. *Games and Economic Behavior*, 61(1):86–112.

Schaner, S. (2018). The persistent power of behavioral change: Long-run impacts of temporary savings subsidies for the poor. *American Economic Journal: Applied Economics*, 10(3):67–100.

Serota, K. B., Levine, T. R., and Boster, F. J. (2010). The prevalence of lying in America: Three studies of self-reported lies. *Human Communication Research*, 36(1):2–25.

Serra-Garcia, M. and Gneezy, U. (2021). Mistakes, overconfidence, and the effect of sharing on detecting lies. *American Economic Review*, 111(10):3160–83.

Shenhav, A., Rand, D. G., and Greene, J. D. (2012). Divine intuition: Cognitive style influences belief in god. *Journal of Experimental Psychology: General*, 141(3):423.

Sims, C. A. (2003). Implications of rational inattention. *Journal of monetary Economics*, 50(3):665–690.

Sobel, J. (2013). Ten possible experiments on communication and deception. *Journal of Economic Behavior & Organization*, 93:408–413.

Speer, S. P., Smidts, A., and Boksem, M. A. (2020). Cognitive control increases honesty in cheaters but cheating in those who are honest. *Proceedings of the National Academy of Sciences*, 117(32):19080–19091.

Speer, S. P., Smidts, A., and Boksem, M. A. (2021). Cognitive control promotes either honesty or dishonesty, depending on one's moral default. *Journal of Neuroscience*, 41(42):8815–8825.

Stagnaro, M. N., Arechar, A. A., and Rand, D. G. (2017). From good institutions to generous citizens: Top-down incentives to cooperate promote subsequent prosociality but not norm enforcement. *Cognition*, 167:212–254.

Verplanken, B. (2006). Beyond frequency: Habit as mental construct. *British Journal of Social Psychology*, 45(3):639–656.

Wang, J. T.-Y., Spezio, M., and Camerer, C. F. (2010). Pinocchio's pupil: using eyetracking and pupil dilation to understand truth telling and deception in sender-receiver games. *American Economic Review*, 100(3):984–1007.

Wood, W., Quinn, J. M., and Kashy, D. A. (2002). Habits in everyday life: Thought, emotion, and action. *Journal of personality and social psychology*, 83(6):1281.

Wood, W. and Rünger, D. (2016). Psychology of habit. *Annual review of psychology*, 67(1):289–314.