Incentives for Conformity and Anticonformity

Fabian Dvorak, Urs Fischbacher and Katrin Schmelz[*]

December 31, 2023

Abstract

While most of the previous literature on social influence focuses on *preferences* for conformity, we provide systematic insights on conformity – as well as anticonformity – in *strategic* environments. Often individuals have to be selected for reward (e.g., promotion) or punishment (e.g., layoffs). To affect the probability of being selected, people might attempt to stand out or to hide in the majority, or to appear as having something in common with the evaluator. We study such strategic incentives for conformity or anticonformity experimentally in three different domains: facts, taste, and creativity. To distinguish conformity and anticonformity from independence, we introduce a new experimental design that allows us to predict participants' independent choices based on transitivity. We find that the prospect of punishment increases conformity, while the prospect of reward reduces it. Anticonformity emerges in the prospect of reward, but only under specific circumstances. Similarity-based selection (i.e., homophily) is much more important for the evaluators' decisions than salience. We also employ a theoretical approach to illustrate strategic key mechanisms of our experimental setting.

Keywords: anticonformity, conformity, homophily, salience, transitivity, evaluation, reward, punishment

JEL Classification: C81, C92, D83, D91

[*] Dvorak: Centre for the Advanced Study of Collective Behaviour, Universitätsstraße 10, 78464 Konstanz, Germany; Department of Environmental Social Sciences, Eawag, Überlandstrasse 133, 8600 Dübendorf, Switzerland, fabian.dvorak@uni-konstanz.de. Fischbacher: Department of Economics, University of Konstanz, Universitätsstraße 10, 78464 Konstanz, Germany; Thurgau Institute of Economics, Hafenstrasse 6, 8280 Kreuzlingen, Switzerland; CESifo, Poschingerstraße 5, 81679 Munich, Germany, urs.fischbacher@uni-konstanz.de. Schmelz: Department of Economics, University of Konstanz, Universitätsstraße 10, 78464 Konstanz, Germany; Santa Fe Institute, 1399 Hyde Park Road, Santa Fe, U.S., katrin@santafe.edu. Support from the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy EXC 2117–422037984 is gratefully acknowledged. We thank Samuel Bowles, Charles Efferson, Armin Falk, Sebastian Fehrler, Simon Gächter, Susanne Goldlücke, Ronald Hübner, Willemien Kets, Kerice Doten-Snitker, Simeon Schudy, Egon Tripoldi, Ro'i Zultan, the research group of the Thurgau Institute of Economics (TWI) and the Department of Economics at the University of Konstanz, as well as participants of various conferences, workshops and seminars for helpful comments and suggestions. We acknowledge the feedback of anonymous referees and the Editor which has helped to improve the paper.

1 Introduction

Conformity and anticonformity are crucial processes contributing to social stability and at the same time promoting diversity on which societal dynamism depends. Various strands of the economic literature draw on the interplay of conformity and anticonformity, for example, research on the exploration-exploitation dilemma in organization theory (Schumpeter, 1934; March, 1991), on rational choice (Simon, 1955), on cultural-institutional evolution (Belloc and Bowles, 2013; Kets and Sandroni, 2021), or on cultural diversity (Kets and Sandroni, 2016). In these studies, conformity and anticonformity are generally assumed to be an intrinsic tendency that is independent of external factors.

People also conform to others' opinions in order to attract reward or avoid punishment (Festinger, 1953; Kelman, 1961; Allen, 1965; Amabile, Goldfarb, and Brackfleld, 1990; Bernheim, 1994; Shalley and Perry-Smith, 2001; Sakha and Grohmann, 2016). Our paper complements this literature by investigating systematically how such external factors affect *strategic* conformity as well as anticonformity. Concretely, we show that the evaluation of individual behavior by peers can provide incentives for conformist or anticonformist *behavior*. For the purpose of this study, we define conformity and anticonformity as behavioral deviations from one's intrinsic choice preference due to information about others' choices. Building on Cialdini and Goldstein (2004), we refer to *conformity* as deviating from one's intrinsic preference *towards* the majority choice, and to *anticonformity* as deviating from one's intrinsic preference *away* from the majority choice.

In this sense, our approach differs from and complements the approach adopted in the fields of biology and anthropology where conformity is considered as taking on the values, preferences and beliefs of others when these are common in a population (e.g., Boyd and Richerson, 1985; Bowles and Choi, 2013; Denton et al., 2020). Instead, our conception relates to Charness, Naef, and Sontuoso (2019, p. 101) who construct "the utility function of a conformist player [as] the sum of a material payoff and a 'psychological bonus' [capturing] the player's intrinsic utility from fitting in." Our core interest is in the part of the utility function determined by the material payoff, and we control for participants' intrinsic inclination to fit in.

Our notion of conformity as a behavioral response to social feedback relates to the economic literature on social learning and imitation (e.g., Apesteguia, Huck, and Oechssler, 2007; Schlag, 1998, 1999; Vega-Redondo, 1997) and information cascades (Banerjee, 1992; Bikhchandani, Hirshleifer, and Welch, 1992; Anderson and Holt, 1997; Guarino, Harmgart, and Huck, 2011). Different from these settings where decisions are typically backward looking, for example when people learn and imitate based on others' payoffs, we investigate settings in which people need to anticipate how others reward or punish and use conformity and anticonformity strategically.

Humans constantly evaluate the actions and intentions of others to decide who deserves punishment and who deserves reward, which is, for example, crucial for the enforcement of social norms (Fehr, Fischbacher, and Gächter, 2002; Fehr and Fischbacher, 2004; Gürerk, Irlenbusch, and Rockenbach, 2006; Guzmán, Rodriguez-Sickert, and Rowthorn, 2007; Henrich et al., 2006; Herrmann, Thöni, and Gächter, 2008). We distinguish between negative evaluation (punishment), where an evaluator selects one out of several individuals for a negative outcome, and positive evaluation (reward), where an evaluator selects one out of several individuals for a positive outcome.

Situations in which an evaluator selects a single individual from a group naturally arise when the implementation of negative and positive outcomes is costly or subject to institutional regulation, as it is common in the labor market and at the workplace. Negative evaluation occurs for example when a team leader has to select one employee for an unpleasant job or, if the company plans layoffs, who will be fired. Typical examples of positive evaluation environments are job interviews or awards.

We study how people strategically respond in anticipation of an evaluators' selection decision. We consider two rules that the evaluators may apply. The first rule is salience, i.e., evaluators' attention may be drawn to the person standing out, determining whom they select. In anticipation, people may attempt to hide behind the majority under the threat of punishment (think about the allocation of an unpleasant task in a team meeting) but try to stand out when a reward is in prospect, as illustrated by Ariel Rubinstein (2013, p. 195f): *"What do you recommend wearing to a job interview? No question, I am the right person to answer this question. I have never given a lecture with a jacket and a tie. I would argue that wearing jeans and a t-shirt is your dominant strategy: If you are a good student, then a department that will not give you a job because of your 'sloppy' appearance does not deserve to have you. If you are mediocre, then there are many other candidates like you and dressing casually is the only way for you to get noticed."* Bordalo, Gennaioli, and Shleifer (2022) review the growing literature on the role of salience in economic decisions.[1] The experimental

[1] Salience also plays an important role in the biological literature, showing that individuals' visual similarity or spatial proximity or may lead to positive or negative consequences. The selfish-herd hypothesis suggests that individuals reduce their risk of dying when forming groups as the risk that a specific individual of the collective is taken by a predator is distributed over all individuals (Hamilton, 1971; Vine, 1971). Studies on fish and crabs show that the average nearest-neighbor distance drops sharply if individuals believe that an immediate threat is present (Viscido and Wethey, 2002; Sosna et al., 2019) and increases if individuals are exposed to food cues (Schaerf, Dillingham, and Ward, 2017). In response to potential benefits, individuals across many species actively express their identity relying on distinctive cues (Tibbetts and Dale, 2007). Standing out has been shown to be crucial for mate attraction, where differentiating oneself from rivals is key to success (Simpson et al., 1999; Buss, 2003).

study on salience in reward and punishment contexts most closely related to our paper is Griskevicius et al. (2006). They find that a self-protection mindset induces conformity, while a mate-attraction mindset induces anticonformity in men (but not in women).[2]

The second potential selection rule is homophily, i.e., the tendency of evaluators to more likely appreciate "those who are alike in some designated respect" (Lazarsfeld and Merton, 1954, p. 23).[3] Responding to homophily would imply to appear similar to the evaluator in order to increase the chances of a reward and avoid punishment. The concept of homophily has widely penetrated the social sciences (see McPherson, Smith-Lovin, and Cook (2001) and Ertug et al. (2022) for surveys), and there is evidence that homophily may benefit individuals who are targets of evaluations. For example, Mäkelä, Björkman, and Ehrnrooth (2010) document a systematic similarity bias in managers' decisions when selecting employees who deserve to be promoted as 'talents'; Opper, Nee, and Brehm (2015) show that homophily increases recruitment chances to China's supreme decision-making body; and similarity between venture capitalists and founders or company executives positively influences funding decisions (Matusik, George, and Heeley, 2008; Hegde and Tumlinson, 2014).

Though studying conformity experimentally has a long tradition across the social sciences, the methodological approach has been challenging. The pioneering literature in psychology has heavily relied on the debated use deception (e.g., Asch, 1951, 1952; Crutchfield, 1955; Hertwig and Ortmann, 2001). An alternative, widely used method has been to present participants the same choice option twice, with and without information about others' decisions (e.g., Griskevicius et al., 2006; Robin, Rusinowska, and Villeval, 2014; Amini et al., 2017), bearing the confounds of a preference for consistency (Falk and Zimmermann, 2017) or an experimenter demand effect.

While the literature on social influence mainly focuses on conformity, studies on anticon-

[2] Griskevicius et al. (2006) first asked their participants to rate how aesthetic they find a series of images. After having been primed towards either self-protection or mate-attraction, participants entered a computer chat with alleged others they thought they would have a face-to-face discussion about aesthetic preferences with later. In the chat room, participants again had to rate one of the previous images, but this time publicly after being informed about the others' alleged ratings (which were pre-programmed to be either uniformly positive or negative).

[3] The term originates from the Ancient Greek *homós* (same, common, similar) and *philía* (love), and we use homophily in its literal sense (unlike a vast stream of literature on homophily that focuses on the formation of ties based on similarity, e.g., Currarini, Jackson, and Pin (2009); Golub and Jackson (2012); Baccara and Yariv (2013); Goldberg and Stein (2018)). Our notion of homophily also closely relates to the literature on similarity-attraction theory (starting with Byrne, 1971), which rests on the idea that people have positive feelings for others who are similar. Far-reaching consequences of this powerful mechanism have been documented in the extensive literature on ingroup favoritism and social identity, typically relying on the minimal group paradigm (as initiated by Tajfel (1970); Turner, Brown, and Tajfel (1979); see Hewstone, Rubin, and Willis (2002) for a review), as well as in the economic literature on taste-based discrimination (Becker, 1971; Riach and Rich, 2002; Bertrand and Duflo, 2017).

formity are scarce (notable exceptions include Ariely and Levav, 2000; Fromkin, 1970; Imhoff and Erb, 2009; Lynn and Harris, 1997; Touboul, 2019) and particularly challenging - not only because anticonformity is rare, but also because it is difficult to disentangle anticonformity from independence (i.e., behavior unaffected by social influence Argyle, 1957; Crutchfield, 1962; Willis, 1963; Willis and Levine, 1976).[4]

Our study introduces a novel experimental technique to identify conformity and anticonformity based on transitivity. We first elicit individuals' preferences in two choices X vs. Y and Y vs. Z in the absence of information about others' choices. Employing transitivity, we then predict the choice of X vs. Z, and we compare the prediction to the individual's actual choice when being informed about others' choices among X vs. Z. Conformity is captured by the frequency of adjustments of choices *towards* the majority choice, and anticonformity by the frequency of adjustments *away from* the majority choice. This technique not only addresses the limitations of earlier methods, but it is crucial to cleanly separate conformity and anticonformity from choices that might appear socially influenced but are actually independent (see Nail, Di Domenico, and MacDonald, 2013, for a discussion).

We investigate the effect of evaluation on conformity and anticonformity in two laboratory experiments, comprising 20 treatments with a total of 871 participants. The general pattern of our implementation is as follows: Participants first make choices without knowing how others decide. Then, they are informed about their group members' decisions and make additional choices. A third party evaluates the group's choices by selecting one choice for reward (resulting in a payoff increase) or punishment (resulting in a payoff deduction), depending on the treatment. We also manipulate the relevance of salience as opposed to homophily in three treatment variations. Treatments without evaluation serve as controls to elicit intrinsic inclinations for conformity and anticonformity in the absence of strategic incentives. We expect the possibility of punishment to induce conformity, and the prospect of reward to limit conformity and induce anticonformity.

Our experiments capture three different choice domains, varying in the degree to which they may foster anticonformity as opposed to conformity. In Experiment 1, participants make simple binary choices in two domains, namely answering questions about objective facts, and expressing their subjective tastes over art paintings. In Experiment 2, we investigate a more complex domain requiring some degree of creativity: participants design several colors and choose which one to be published. We expect increasing anticonformity and decreasing

[4] Nonetheless, anticonformity can be a vital mechanism, for example for opinion dynamics. A single anticonformist response can break unanimity, which is a particularly strong determinant of conformity in subsequent choices (Asch, 1955), and can prevent information cascades or influence the polarization of opinions (Juul and Porter, 2019; Siedlecki, Szwabinski, and Tomasz, 2016).

conformity across those three domains, particularly in the reward treatments.

The central result of the experiments is that people show strategic conformity to avoid being punished, and they show reduced conformity when facing rewards. Strategic anticonformity to attract a reward is rare and occurs only under certain conditions.[5]

This is well in line with the evalutors' behavior, creating incentives for conformity if the consequence for the selected individual is punishment, and under certain reward conditions creating incentives for anticonformity. Concerning the mechanisms, evaluators' selection decisions are mainly driven by homophily. Salience plays a minor role and turns out to be relevant only in treatments where this mechanism is made very salient to participants.

Across our domains, conformity decreases and anticonformity increases from objective facts over arts taste to creativity.[6] Moreover, we observe individual heterogeneity in strategic conformity and anticonformity.[7]

We complement the laboratory experiments with a theoretical framework to illustrate how punishment and reward affect conformity and anticonformity under the salience and the homophily rule. Our model relies on the setting of Experiment 1, capturing binary choices subject to evaluation. The model predicts that if evaluation is based on salience, punishment incentivizes conformity and reward incentivizes anticonformity. If evaluation is based on homophily, again, punishment incentivizes conformity. Reward leads to less conformity than punishment and can even evoke anticonformity. The model shows that evaluation has the potential to induce strategic conformity and anticonformity in the sense that both can be rational responses to social influence in order to avoid punishment and attract reward.

[5] Our finding of ample conformity and rare anticonformity is in line with the evolutionary literature showing that copying the behavior of others is a superior strategy (Rendell et al., 2010), while deviating from one's group can threaten group membership and lead to ostracism (Mahdi, 1986; Boehm, 1993, 2000; Wiessner, 2002; Boyd, Gintis, and Bowles, 2010).

[6] The observation that of highest conformity when answering knowledge questions is consistent with the extensive literature on social learning (e.g., Banerjee, 1992; Bikhchandani, Hirshleifer, and Welch, 1992; Lee, 1993, 1998; Anderson and Holt, 1997; Vives, 1997; Smith and Sorensen, 2000; Banerjee and Fudenberg, 2004). We also find conformity in arts tastes (including in the *Reward* treatment), which is in line with the literature on frequency dependent social learning (Boyd and Richerson, 1982, 1985; Efferson et al., 2008; McElreath et al., 2008).

[7] Our finding on heterogeneity in strategic conformity and anticonformity is in line with studies on heterogeneity in preferences for conformity or anticonformity (Argyle, 1957; Brehm, 1966; Corazzini and Greiner, 2007; Fatas, Heap, and Arjona, 2018; Goeree and Yariv, 2015; Jones, 1984; Jones and Linardi, 2014; Wright, London, and Waechter, 2009), which have been related to both individual traits (Ariely and Levav, 2000; Fromkin, 1970; Imhoff and Erb, 2009; Lynn and Harris, 1997) and cultural variation (Bond and Smith, 1996; Cialdini et al., 1999; Kim and Markus, 1999; Yamagishi, Hashimoto, and Schug, 2008).

2 Theory

Our theoretical model illustrates the incentive structure of our Experiment 1.[8] The model shows how reward and punishment affect conformity and anticonformity, predicting that under reward there is less conformity than under punishment, and reward may even induce anticonformity. We consider two decision rules for the evaluator, salience and homophily.[9]

In our setup, an evaluator observes the binary choices of a group of individuals and selects one individual from the group for reward or punishment. If the evaluation is based on salience, the evaluator will select the person who stands out. If the evaluation is based on homophily, the evaluator will select a person whose choice matches the evaluator's taste. For simplicity, we assume that only one of the two rules applies and treat the two cases independently.

If the evaluators apply the salience rule, the model predicts conformity in case of punishment and anticonformity in case of reward. The intuition is as follows. Conformity avoids to be singled out, which is good when the singled-out person is punished. Anticonformity ensures to be the singled-out person, which is good when this person is rewarded.

If the evaluation is based on homophily, the player must assume that the evaluator's taste is likely to match the majority's choice. In the case of punishment, conforming to the majority is optimal for two reasons: First, the evaluator is more likely to punish the minority choice, and second, joining the majority means that there are more people with the same choice - which protects against being selected. The case of homophily-based reward is more complex. An argument for blending is that the evaluator is more likely to want to reward the majority choice. At the same time, the probability of receiving the reward is shared among all participants with the same choice - which argues for standing out. Thus, optimal behavior depends on the correlation structure of preferences in the population.

Model

There are $N \geq 3$ group members, consisting of players A_1, ... A_{N-1} and a player B, and as well as a group-external evaluator E. The group members make a binary choice between two options X and Y. All players prefer one of the options. The group members derive a utility $\tau_i > 0, i \in 1..N$ if they choose the option that corresponds to their own taste. The evaluators are not modeled based on maximizing utility. They just implement a decision rule. We assume that the players' preferences are correlated in the following way. There

[8] Experiment 2 features a more complex choice setting that goes beyond the simple setup used for the theoretical model, especially as it allows for more than two choice alternatives.

[9] We also discuss a rule based on performance in Appendix A for interested readers. This rule is secondary as it is not in the core interest of our study and applies only to specific environments in our setting.

is a generally preferred option. The probability for each specific option to be generally preferred is $\frac{1}{2}$. Each player independently prefers this option with probability $p > \frac{1}{2}$.[10] A high probability p means that the preferences in the population of players are similar. A low probability p, i.e. a probability close to $\frac{1}{2}$, means that the preferences in the population of players are mixed.

First, the A players (i.e. players A_1, ... A_{N-1}) decide. They make their choices simultaneously. Player B chooses after observing the choices of the A players. The evaluator E selects one of the N group members without being informed of who player B is. In the punishment treatment, the selected player receives a monetary deduction of m. In the reward treatment, the selected player receives a monetary payment of m.

The utility of the players A and B consists of the utility from the money m if they are selected, and of the utility τ_i if they choose the option that corresponds to their taste. We assume that the components are additive and that τ_i is expressed in monetary terms. So, the utility equals $M + \tau_i$, where $M = 0$ if the player is not selected, $M = m$ if the player is rewarded, and $M = -m$ if the player is punished. The cumulative distribution function T of τ_i is common knowledge, and we assume that it is continuous and strictly increasing between 0 and a value τ_{max}.

The evaluator decides according to a rule. The salience-based rule means that he selects someone from the minority if there is one. Otherwise he chooses a player randomly. The homophily-based rule means that, if possible, the evaluator rewards someone who has chosen in accordance with the evaluator's own taste and punishes someone who has chosen against the evaluator's taste.

We describe the equilibria for the case of $N = 3$. We give some insights on the general case of $N > 3$ in Appendix A. The A players have two strategies: following their own taste and switching (i.e., choosing opposite to their own taste). Since player B is not informed about the identity of the A players, B can only condition on the number of A players who decide according to B's taste. Thus, player B has eight pure strategies.

Responses to evaluation based on salience

Proposition 1 (Salience-based punishment)**.** *The A players follow their own taste. If the A players both disagree with B then B chooses as the A players if $\frac{\tau_B}{m} < \frac{2}{3}$, and B is indifferent if $\frac{\tau_B}{m} = \frac{2}{3}$. In all other cases, B follows his own taste.*

Proposition 2 (Salience-based reward)**.** *There is a unique symmetric equilibrium, which*

[10] This means that the probability that two players prefer the same option is $\sigma = p^2 + (1-p)^2$. Conversely, we can calculate p based on σ: $p = \frac{1}{2} + \frac{1}{2}\sqrt{2\sigma - 1}$.

is characterized as follows. The A players choose according to their own taste if $\frac{\tau_A}{m} > K$ where K is a constant that depends on T, and p. They choose against their preferred option if $\frac{\tau_A}{m} < K$, and are indifferent if $\frac{\tau_A}{m} = K$. If the A players both agree with B then B chooses contrary to the A players if $\frac{\tau_B}{m} < \frac{2}{3}$, and is indifferent if $\frac{\tau_B}{m} = \frac{2}{3}$. In all other cases, B chooses according to his own taste.

The optimal behavior of player B follows directly from the definition of salience-based evaluation. Concerning the behavior of A, it is intuitively clear that if there is an incentive for conformity, then A players should coordinate, which they best achieve by following their own taste. In the case of reward, the A players may have an incentive to deviate from their own taste in order to make it more difficult for B to stand out. A formal proof of the propositions and a discussion of the symmetry assumption are provided in Appendix A.

Responses to evaluation based on homophily

We now consider responses to social evaluation based on homophily. We start with some terminology on player B's decision: It is called *independent* if it coincides with player B's own taste. It is called *conformist* if, in case of disagreement with the A players, B neglects his own taste and follows the choice of the majority. It is called *anticonformist* if, in case of agreement with both A players, B neglects his own taste and makes a minority choice.

We derive the following propositions (the proofs are provided in Appendix A).

Proposition 3 (Homophily-based punishment)**.** *Independent of the strategy of player B, the A players always follow their own taste. B is conformist if $\frac{\tau_B}{m} < 2(p - \frac{1}{2})^2 + \frac{1}{6}$, otherwise B is independent. (In case of equality B is indifferent between conformity and independence.)*

Proposition 4 (Homophily-based reward)**.** *Independent of the strategy of player B, the A players always follow their own taste. B is conformist if $\frac{\tau_B}{m} < \frac{1}{3} - 2p(1-p)$. B is anticonformist if $\frac{\tau_B}{m} < \frac{2}{3} - \frac{p^4+(1-p)^4}{p^3+(1-p)^3}$. (In case of equality B is indifferent between conformity or anticonformity and independence.)*

Figure 1 shows the limits for conformity and anticonformity in the two treatments. Conformity is more likely in the case of punishment and anticonformity is only possible in the case of reward. Under punishment, conformity can exist for any values of p because $2(p - \frac{1}{2})^2 + \frac{1}{6} > 0$. Under reward, conformity can only exist if $\frac{1}{3} - 2p(1-p) \geq 0$, which is the case if $p \geq \frac{1}{2} + \frac{\sqrt{3}}{6} \approx 0.789$. Anticonformity can only exist if $\frac{2}{3} - \frac{p^4+(1-p)^4}{p^3+(1-p)^3} \geq 0$, which is the case if $p \leq \frac{1}{6}(3 + \sqrt{6\sqrt{3} - 9}) \approx 0.697$.

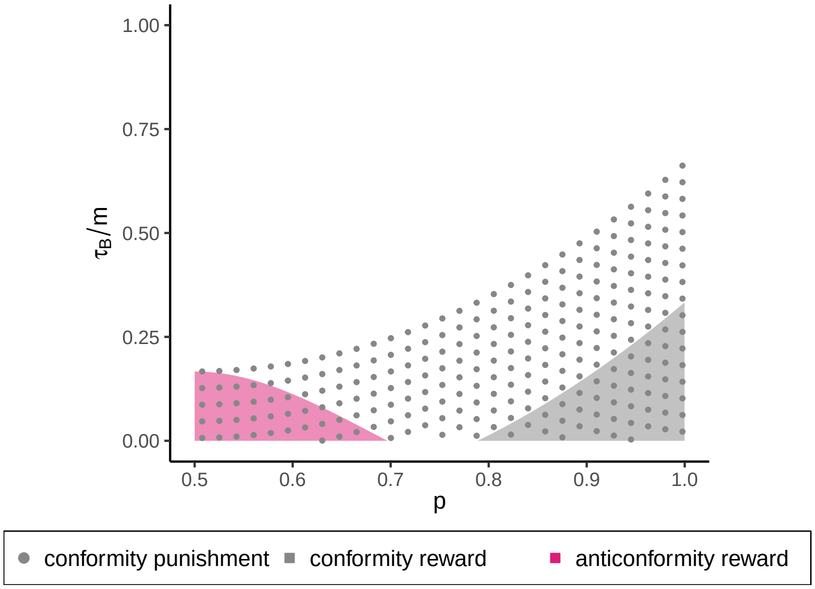


Figure 1: Limits for conformity and anticonformity in the case of homophily-based punishment and reward. The areas show the combinations of p and $\frac{\tau_B}{m}$ for which there is conformity or anticonformity in the respective treatments.

3 Experiment 1

We study the effect of evaluation on conformity and anticonformity in laboratory experiments. The general pattern of our implementation is as follows: In groups of three, two group members make individual choices between two options. The third group member learns about the choices of the other two members and then also makes a choice between these two options. A third party evaluates the group's three choices by selecting the choice of one person for reward or punishment, depending the treatment. We also implement control treatments without evaluation to elicit participants' intrinsic inclinations for conformity and anticonformity in the absence of strategic incentives.

We investigate strategic conformity and anticonformity in various settings which are determined by three dimensions: choice domains (i.e., *Facts* and *Taste*), incentives (i.e., *Reward*, *Punishment* and *Control* treatments), and the importance of salience-based evaluation. An overview of our experimental setup is provided in Table 1, and the three dimensions are explained in the remainder of this section.

Our main interest is in strategic conformity and anticonformity, as captured by the *Reward* and *Punishment* treatments. They serve to measure whether and to what extent positive and negative incentives affect behavioral adjustments towards or away from the majority, as detailed in Section 2. The purpose of the domains is to explore how these incentives interact with the objective (*Facts*) and subjective (*Taste*) nature of the choice environment. Finally, given the low relevance of salience as compared to homophily determining evaluators' choices in our initial treatments (labeled *S0*), we add treatments *S1* and *S2* inducing more salience-based evaluations to conclusively understand whether or not this mechanism is relevant in such settings.

3.1 Choice domains: facts and taste

Experiment 1 captures two choice domains, differing in the degree to which they may foster anticonformity as opposed to conformity: *Facts* and *Taste*. In these two domains, participants always select one out of two options.

In the domain of *Facts*, participants face a series of difficult factual questions which have an objectively correct answer, though the data underlying the answers are very similar for both options and beyond the general knowledge of typical university students.[11] The

[11] Examples are: "Which country is older: Ghana or Niger?" Ghana was founded in 1957, one year before Niger was founded in 1958, and is therefore the correct answer. "Who has sold more records in Germany: Britney Spears or Bon Jovi?" At the time of data collection, Bon Jovi was with 5,150,000 records slightly ahead of Britney Spears with 5,050,000 records. "Which airport had more passengers in 2014: Aeropuerto Madrid Barajas or Miami International Airport?" Madrid was somewhat more busy with 41,822,863

Table 1: Overview of the setup of Experiment 1

Domains (within-subjects)	Facts & Taste							
Incentives (between-subjects)	Reward			Punishment			Control	
Salience (between-subjects)	S0	S1	S2	S0	S1	S2	S0	S2
Pre-stage: salience training Payment for successful coordination on separate choice sets.	-	-	✓	-	-	✓	-	✓
Stage 1: uninformed choice Participants go through 20 binary choice sets in the absence of information about others‘ choices.	✓	✓	✓	✓	✓	✓	✓	✓
Stage 2: informed choice Participants go through 10 binary choice sets, knowing how their group members have decided in this situation.	✓	✓	✓	✓	✓	✓	✓	✓
Stage 3: evaluation The evaluator selects one of the group‘s choices.	✓	✓	✓	✓	✓	✓	-	-
Group members’ incentives								
Selected group member is rewarded.	✓	✓	✓	-	-	-	-	-
Selected group member is punished.	-	-	-	✓	✓	✓	-	-
Evaluator’s (salience) incentives								
Evaluators are paid for successful coordination with other evaluators.	-	✓	✓	-	✓	✓	-	-

objective, fact-based nature of this domain may allow for social learning, and the difficulty gives room for conformity to emerge.

In the *Taste* domain, participants choose among two similar art paintings.[12] As this domain involves subjective art preferences rather than objective facts, there is no right or wrong answer, and we therefore expect less conformity compared to the *Facts* domain. Note that, based on our model as explained in Section 2, we nevertheless expect some degree of strategic anticonformity in the *Facts* domain under *Reward* incentives.

3.2 Eliciting conformity and anticonformity based on transitivity

We measure conformity by the frequency of adjustments of participants' choices *towards* the majority choice, and anticonformity by the frequency of adjustments *away from* the majority choice. Participants first make choices without knowing how others decide. Then, they are informed about others' decisions and choose again. We measure conformity and anticonformity by comparing the choices without and with information about others' choices.

Adjustments occur if participants deviate from their intrinsic preference that we predict based on transitivity. To predict intrinsic preferences based on transitivity, we form groups of three participants and use triplets of choice alternatives. Each triplet (X, Y, Z) generates three binary choice situations (X vs. Y, Y vs. Z, and X vs. Z).

Out of these three possible binary choices, in Stage 1, each group member faces two binary choice situations without being informed of the choices of their group members (referred to as uninformed choices). These binary choice situations are assigned to the group members such that the remaining binary choice situation is different for each group member. For example, if participant 1 (referred to as P1) faces the option pairs X vs. Y and Y vs. Z, P2 faces X vs. Y and X vs. Z, and P3 faces X vs. Z and Y vs. Z.

In Stage 2, participants decide in the remaining binary choice situation they have not faced yet (P1: X vs. Z; P2: Y vs. Z; P3: X vs. Y). Before making their choice in Stage 2, participants are informed about the Stage 1 decisions of the two other group members in the same binary choice situation. In our example, before choosing between X and Z in Stage 2, P1 is informed about P2's and P3's decisions in the choice between X and Z. Thus, we refer to the Stage 2 decisions as informed choices.

To detect conformity and anticonformity, i.e., deviations from a participant's intrinsic preference towards or away from the majority, we predict a participant's choice in Stage 2

passengers, compared to Miami with 40,941,879 passengers.

[12] For example, two variants of the *Garden of the Artist* by Monet, two variants of a bride and a groom by Chagall, or flowers in a vase by van Gogh and Renoir. Full lists of the factual questions and paintings used in Experiment 1 are provided in Appendix C.

assuming transitivity over the three choice alternatives based on the two uninformed choices in Stage 1, and compare it to the actual informed choice in Stage 2. We also elicit the strength of participants' preferences in these choices on a continuous scale (as shown in Appendix C): after each uninformed choice, participants are asked how much they prefer their selected item over the alternative. This measure not only serves to better understand intransitivity, but also to predict the uninformed choices in cases where in Stage 1, a participant prefers X over Y as well as X over Z.

For each of the *Facts* and *Taste* domains, we use 10 sets of triplets per group. Accordingly, Stage 1 consists of 20 uninformed binary choice scenarios without information about others' choices, and another 10 informed binary choice scenarios in Stage 2, where participants can condition their choices on their group members' selections.

To provide an incentive for selecting according to participants' actual arts preferences, participants received one of their chosen paintings in the form of an art postcard at the end of the session. For each group within a given session, a different set of postcards was randomly selected for being handed over, such that choosing a unique postcard would also apply in the context of the entire session. To provide some incentive for answering the factual questions correctly, participants were shown the correct solutions along with their answers once all choices in the *Facts* domain were made.

Discussion on transitivity: merits and limitations

Measuring conformity (and, to a lesser extent, anticonformity) experimentally has a long tradition across the social sciences, starting with Asch (1951). Typically, these phenomena have been studied by instructing confederates to give false answers to questions that do have an obviously correct answer, inducing a sharp mismatch between the subject's objective observations and the judgments of the pre-instructed group of confederates (as in Asch's experiments); or by using other forms of deception like mimicking others' responses by an apparatus (e.g., Crutchfield, 1955). For good reasons, the use of deception is meanwhile heavily debated, particularly in economics (Hertwig and Ortmann, 2001).

An alternative approach has been to present participants the same choice option twice, with and without information about others' decisions (e.g., Griskevicius et al., 2006; Robin, Rusinowska, and Villeval, 2014; Amini et al., 2017). Facing the same choice set twice implies that participants are likely to remember their uninformed first choice, which is often intended. This carries two potential confounds. First, people may have a preference for consistency (see Falk and Zimmermann, 2017) and thus stick to the same option in their informed second choice, which may reduce the observed effect of social influence. Second, asking the same question twice may trigger an experimenter demand effect that can go in either direction.

Our technique to measure conformity and anticonformity using transitivity mitigates these concerns. In our setting, participants never face the same pairs of options twice. Even though our design does not fully exclude the possibility that participants may remember their uninformed choices, they would have to triangulate the informed choice that would be consistent with their two uninformed choices in order to make a deliberate consistent choice, or to respond to a perceived experimenter's demand. Given that participants go through a series of 20 uninformed choices, remembering each pair as a basis for triangulating is unlikely. If participants attempted to be consistent by remembering all choices they made in Stage 1 and triangulating their choices in Stage 2, the deviations we observe in the experiment would reflect lower bounds of conformity and anticonformity.

The design of Experiment 1 accounts for violations of transitivity (Tversky, 1969; Loomes, Starmer, and Sugden, 1991). Our measures of conformity and anticonformity – estimated in situations where a participant faces identical choices by the other group members in Stage 2 (referred to as majority information) – are corrected for baseline intransitivity – estimated from situations in Stage 2 where a participant faces two different choices of the other two group members (referred to as mixed information).

The transitivity approach does not work reliably when subjects (strategically) misreport their uninformed choice. In fact, according to our theoretical model, this can be the case in the *Reward* treatments if the evaluation is based on salience. Since players should shift to their actually preferred choice in the mixed information scenario in Stage 2, we are able to assess this potential problem by comparing the intransitivity levels across the treatments and scenarios (see Section 5).

In a nutshell, our new approach to measure conformity and anticonformity based on transitivity reduces the potential confounds of consistency and an experimenter demand effect, and it is robust to violations of the transitivity assumption.

3.3 Reward and punishment treatments

The experimental setup described above reflects our *Control* treatment (C), where participants' choices are disclosed to the other group members, without any monetary incentives involved. To study strategic incentives for conformity and anticonformity, we implement *Reward* (R) and *Punishment* (P) treatments, and we study how these positive and negative incentives interact with the two domains (i.e., *Facts* and *Taste*).

The general principle of the *Reward* and *Punishment* treatments is that one of the three group members is assigned a bonus pay or a deduction in Stage 3. Using a within-subjects design, after having made their informed choices in Stage 2, each participant takes the role of an evaluator and evaluates the choices of another group. Evaluators are shown the three group members' chosen options of a binary choice set (e.g., X vs. Y). Thus, they either see three copies of the same option (e.g., X, X, X), or they see one option twice and the other option once (e.g., X, X, Y). These three chosen options are composed of two uninformed choices from Stage 1 and one informed choice from Stage 2, and the evaluators do not know which of these options are the result of an uninformed or informed choice. They select one of the three displayed choices (which are shown in a randomized order on each evaluator's screen) and the corresponding group member receives a reward or punishment, depending on the treatment. See Appendix C for an example.

Each evaluator makes 30 such decisions in Stage 3. The evaluation situations evaluators face are also derived from triplets they have faced themselves when deciding in the role of a group member. This allows us to investigate the relevance of homophily for the evaluators' decisions.

In the *Reward* treatments, 10 euros are added to the final payoff of the selected participant, while in the *Punishment* treatments, 10 euros are deducted from the selected participant. Participants receive a flat payment of 16 euros in the *Control*, 30 euros in the *Punishment*, and 20 euros in the *Reward* treatments. The average payoff in the *Control* treatment is lower because these sessions were shorter as they did not include Stage 3, such that the treatments are comparable in their hourly payment. The flat payments are higher in the *Punishment* than in the *Reward* treatments to ensure equivalent lowest payoffs (20 euros in both cases) as well as similar average earnings. The fact that the *Reward* treatment incurs gains whereas the *Punishment* treatment incurs losses is an inherent feature of our design.[13]

For each domain, one evaluation decision per group was randomly selected for payment.[14] The *Reward* and *Punishment* treatments were implemented in a between-subjects design, whereas the *Facts* and *Taste* domains were implemented within-subjects. The order of the two domains was balanced across the experimental sessions. At the end of the three stages of each domain, participants received feedback about the evaluation decisions in their group and their own payoffs.

[13] Our main focus is on how reward and punishment affect (anti)conformity compared to an environment in the absence of such features, and we therefore implemented this more natural version of a control treatment (instead of having separate control treatments where reward and punishment would be allocated randomly at the end of a session).

[14] The main reason for paying only one instead of all choices is the experimental feature that participants receive a physical copy of their selected arts picture, and they should have the possibility to be unique in their entire session.

3.4 Salience treatments

As explained in Section 2, two potential mechanisms driving evaluators' decisions are homophily and salience. In the treatments presented so far, to our surprise, the data show little evidence for salience-based evaluations. We therefore implemented two more treatment variations pushing the mechanism of salience. We study the relevance of salience in three treatment variations, increasing the weight of salience as opposed to homophily. These salience treatments apply in the same way to the *Punishment* and *Reward* treatments.

Evaluators are not incentivized for their evaluation decisions in the *S0* treatments (as described above in Subsection 3.3). Applying coordination as a standard method to study salience (Mehta, Starmer, and Sugden, 1994a,b), in the *S1* and *S2* treatments, evaluators are incentivized to coordinate their evaluation decision with other evaluators. Several participants evaluate the same choices (as in *S0*), but now, their payoff increases with each other evaluator who selects the same participant. An evaluator's payoff increases by 0.002 euros for each percentage point of the total number of other evaluation decisions in the same experimental session that correspond to this decision. As there is no communication involved, coordination of the evaluation decisions has to be achieved by selecting the choice that is generally considered as salient.

In the *S2* treatments, we further induce salience-based evaluation by implementing a coordination training stage before the actual experiment starts. Participants are incentivized to coordinate their choices in a separate set of items (different from the actual experiment, these items do not reflect choices by other participants). In this training stage, participants are shown sets of three icons, where either all three icons are exactly the same, or one icon is different from the other two.[15] In each set, they are asked to select one icon, and their payoff increases by 0.002 euros for each percentage point of the total number of other participants in same experimental session that correspond to their decision. As in the later coordination tasks of Stage 3, the position of the icons on the evaluators' screens is randomized to rule out the possibility of location-based coordination.[16]

3.5 Procedures

We conducted a total of 16 experimental sessions (six sessions of the *S0* treatments and four sessions of the *S1* treatments in 2017/2018, and another 6 sessions of the *S2* treatments

[15] For example, three copies of an icon showing one dot, or two copies of a one-dot-icon and one three-dots-icon. The full list of training icons is provided in Table 16 of the Appendix.

[16] Note that we also include a *S2* version of the *Control* treatment to see whether the coordination training has any effect on participants' choices in the first two stages. Obviously, we do not have a *S1 Control* treatment because there are no evaluators involved who could coordinate in Stage 3.

in 2023) with 396 students at the University of Konstanz in Germany. Participants were recruited via ORSEE (Greiner, 2015) in the earlier sessions and via hroot (Bock, Baetge, and Nicklisch, 2014) in the later sessions. The experiments were conducted with z-Tree (Fischbacher, 2007), and instructions were shown as PDFs on participants' screens using E-nstructions (Schmelz, 2011). The mean age of the participants was 22.3 years, and 54% were female. Table 2 summarizes the number of sessions, participants, groups and choices in each experimental condition.

Table 2: Summary of treatment data in Experiment 1

Domain	Facts								Taste							
Incentive	Reward			Punishment			No		Reward			Punishment			No	
Salience	S0	S1	S2	S0	S1	S2	S0	S2	S0	S1	S2	S0	S1	S2	S0	S2
Sessions	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
Participants	54	51	45	54	51	39	60	42	54	51	45	54	51	39	60	42
Groups	18	17	15	18	17	13	20	14	18	17	15	18	17	13	20	14
Informed choices	540	510	450	540	510	390	600	420	540	510	450	540	510	390	600	420
Evaluations	1,620	1,530	1,350	1,620	1,530	1,170	1,800	1,260	1,620	1,530	1,350	1,620	1530	1,170	1,800	1,260

The number of participants reflects the number of statistically independent observations in Experiment 1. All participants take the role of an evaluator and evaluations were elicited using the strategy method (Selten, 1967), which leads to the relatively large number of evaluation decisions.

4 Experiment 2

As shown in the literature discussed earlier, conformity has been documented to be a strong force in human interactions, while anticonformity is much more rare – and this is also what we observe in Experiment 1. However, encouraging uniqueness is an important goal in the literature on organizations, and creativity has been linked to the degree of (non-)conformity in a society. For example, Shane (1992) shows that individualistic countries are more inventive than more conformist societies. Similarly, according to Goncalo and Staw (2006), individualistic rather than collectivist values foster creativity.

Experiment 2 enters the domain of *Creativity* and serves to create conditions inviting anticonformity to be potentially expressed (more). Because creativity cannot be studied in a binary setting, it employs a substantially different and more complex setup than Experiment 1, abandoning the setting of choosing between binary options, also departing from the binary nature of our theoretical model and from inferring participants' preferences based on transitivity.

To measure conformity and anticonformity in a creativity task, we need to quantify how close or far two outcomes of a creative process are. To do so, we develop a new design making use of the RGB color space. Participants design several colors and choose one of those multiple self-designed options to be displayed to others. Evaluators can reward or punish design choices, and we also include coordination treatments to foster salience-based evaluations.

We expect the color creation task to induce more anticonformity compared to the domains in Experiment 1 because designers exert an activity which is likely to be intrinsically motivating, they may identify with their creative output more than with a selection in a predefined choice set, and they have the possibility to express their identity and differentiate themselves from others by opting for a more unique color.

Experiment 2 consists of four treatments following a 2 x 2 design, implemented between-subjects, as shown in Table 3. Again, the main design dimension captures the incentives (*Reward* and *Punishment*), and we also increase the importance of salience-driven evaluation from the *S0* treatments (eliciting baseline salience-based evaluations) to the *S1* treatments (adding coordination incentives for evaluators).

4.1 Choice domain: creativity

To let participants design colors, we developed a color generation interface where each designer starts out with the same eight colors (red, green, blue, yellow, magenta, cyan, black and white), representing the vertices of the three-dimensional RGB color space. Participants can then generate new colors by average mixing of two colors. Newly generated colors can be stored and reused to generate further colors. By repeatedly executing these steps, every color in the RGB color space can be approximated.[17]

Designers have two minutes at the beginning of each of the eight rounds of the experiment to create new colors by mixing preexisting colors. Their created colors from past rounds are available to them in future rounds. During this color creation phase, designers can make a short list with up to four of their created colors.

4.2 Eliciting conformity and anticonformity based on adjustments

The basic procedure to elicit deviations from intrinsic preferences towards and away from the majority in Experiment 2 is to compare pre-selected options in the absence of information

[17] A picture of the color generation interface is shown in Appendix C, and a movie illustrating the procedure of generating colors is provided under https://fdvorak.com/videos/creativity-task.mp4.

Table 3: Overview of treatments in Experiment 2

Domain	Creativity			
Incentives (between-subjects)	Reward		Punishment	
Salience (between-subjects)	S0	S1	S0	S1
Pre-stage: color generation All participants design colors for two minutes.	✓	✓	✓	✓
Stage 1: uninformed choice Designers submit their pre-selected color to their group of designers in the absence of information about others‘ choices.	✓	✓	✓	✓
Stage 2: informed choice Designers can adjust their selected color, knowing their group members’ pre-selected colors.	✓	✓	✓	✓
Stage 3: evaluation The evaluator selects one of the group‘s choices.	✓	✓	✓	✓
Group members’ incentives				
Selected group member is rewarded.	✓	✓	-	-
Selected group member is punished.	-	-	✓	✓
Evaluator’s (salience) incentives				
Evaluators are paid for successful coordination with another evaluator.	-	✓	-	✓

Note: These stages are repeated in each of the eight rounds of the experiment.

about others' choices with their adjustments following information about others' choices.

In groups of four, each designer first creates a private shortlist of their self-created colors (pre-stage color generation), and then pre-selects one color to be published within their group in Stage 1 (referred to as uninformed choice). In Stage 2, after having seen the four preselected colors of the group, each designer has the opportunity to replace their pre-selected color by a different color from their short list (referred to as informed choice). The final published color set of a group consists of the uninformed Stage 1 choices of three designers and the informed Stage 2 choice of one randomly selected designer.

As designers generate their own colors, the choice alternatives differ across designers (different from Experiment 1 where at least two out of three group members' choices are identical by design). Accordingly, conformity (resp. anticonformity) is captured by a color similar (resp. dissimilar) to the colors of the other designers. To quantify the similarity of a color to the three pre-selected colors of the other group members, we use various measures based on Euclidean distance (two different color spaces and three ways to aggregate the distance to the other three subjects, see Appendix B for details).

An adjustment occurs when the informed choice differs from the uninformed choice. We consider a decision as conformist if the informed choice is closer to the colors of the other group members than the uninformed choice; and we consider a decision as anticonformist if the informed choice is further away from the colors of the other group members than the uninformed choice.[18]

4.3 Reward and punishment treatments

In Stage 3, the uninformed choices of three designers and the informed choice of one randomly selected designer are transmitted to an evaluator, who does not know which of the four colors is the informed choice and which are uninformed color submissions. The evaluator selects one of those four colors, and the corresponding designer receives a bonus (*Reward* treatment) or a deduction (*Punishment* treatment).

To implement reward and punishment, respectively, 2 euros are added and deducted, respectively, from the payoff of the selected designer. To ensure similar average earnings across treatments, designers receive a flat payment of 20 euros in the *Punishment* and 12 euros in the *Reward* treatments. Evaluators receive a flat payment of 16 euros. After each round, all designers receive feedback about the evaluation decision in their group, yielding the possibility to converge or diverge as a group over the course of the experiment.[19]

[18] Obviously, predicting participants' intrinsic preferences based on transitivity is not possible for the multinominal choices of Experiment 2.

[19] We did not include a *Control* treatment without incentives in Experiment 2, as it is unclear what an

4.4 Salience treatments

The experimental setup described so far reflects the *S0* treatments of Experiment 2. We also include *S1* treatments with coordination incentives, where two evaluators are assigned to the same group (instead of only one evaluator in *S0*), such that coordination is possible. Both evaluators receive an additional payment of 2 euros if their decisions coincide, and the decision of one randomly selected evaluator is implemented to determine the designers' payment of a given round.[20]

In Experiment 2, participants take the fixed roles of a *designer* or an *evaluator*. Evaluators are tied to a given group over the eight rounds to avoid spillovers from having evaluated other groups, which might bias their salience perceptions given the multidimensional nature of options. The same concern would apply had evaluators participated as designers themselves. To nevertheless let evaluators gain experience with the designers' setting, they participated in the pre-stage of each round, where they could generate colors just for play.

4.5 Procedures

Each session consisted of three parts. After having performed the experimental treatments as detailed above in the first part, we elicited which color is generally considered to be salient by performing a Krupka-Weber coordination task (Krupka and Weber, 2013) across the four colors shown to evaluators in the second part. Finally, in the third part, participants rated how beautiful and interesting they find each of those four colors. We used two continuous scales ranging from zero (not beautiful / not interesting at all) to one (very beautiful / very interesting).

The experiment was conducted in 16 experimental sessions with 475 students at the University of Konstanz in 2018. Participants were recruited via ORSEE (Greiner, 2015), excluding participants who had participated in Experiment 1. The experiment was implemented in z-Tree (Fischbacher, 2007), and instructions were shown as PDFs on participants' screens using E-nstructions (Schmelz, 2011). The mean age of the participants was 21.3 years, and 63% were female. Table 4 summarizes the numbers of sessions, participants, groups and choices in each treatment.

appropriate *Control* treatment would be. Eliminating Stage 3 and going through Stages 1 and 2 without the outcomes being shown to an evaluator may feel odd and confusing to participants as there would be no purpose in doing so. Showing the outcomes to an evaluator in Stage 3 without implementing monetary incentives would lean towards the *Reward* treatment as the evaluator pays attention to the selected colors.

[20] When conducting the *S0* treatments, we did not anticipate to observe so little salience-based evaluation, and that we would add treatments pushing this mechanisms. So, it seemed natural to assign one evaluator to each group. Even though the designers' decisions are shown to one evaluator in *S0* but to two evaluators in *S1*, designers' monetary incentives remain unchanged across these treatment variations.

Table 4: Summary of treatment data of Experiment 2

Domain	Creativity			
Incentive	Reward		Punishment	
Salience	S0	S1	S0	S1
Sessions	4	4	4	4
Participants	115	120	120	120
Groups	23	20	24	20
Informed choices	736	640	768	640
Evaluations	184	320	192	320

The number of groups reflects the number of statistically independent observations in Experiment 2.

5 Experimental results

In the first part of this section, we show our main results on responses to incentives for conformity and anticonformity, and how these responses interact with our experimental settings. We then turn to our findings on the importance of the homophily and salience rules driving the evaluators' decisions, which determine the actual incentives for conformity and anticonformity in our data. Throughout this section, we present the results of our two experiments jointly.

5.1 Responses to incentives for conformity and anticonformity

To depict the responses to incentives in our treatments, we rely on the conceptual framework provided by the Willis-Nail model of social response (Willis, 1965; Willis and Levine, 1976; Nail, 1986; Nail and Van Leeuwen, 1993; Nyczka and Sznajd-Weron, 2013; Nyczka et al., 2018). The three vertices of the model space represent the three canonical responses to social influence: conformity (C), anticonformity (A), and independence (I).

Average responses to incentives

Figure 2 shows the average responses to incentives in our treatments according to this conceptual framework. We operationalize the horizontal independence dimension as the relative frequency of adjustments of the informed choices in either direction, i.e., the sum of the relative frequencies of adjustments towards and away from the majority. The vertical conformity-anticonformity dimension (vertices C and A) captures the net direction of the adjustments, i.e., the relative frequency of adjustments of the informed choice towards the majority minus the relative frequency of adjustments away from the majority.

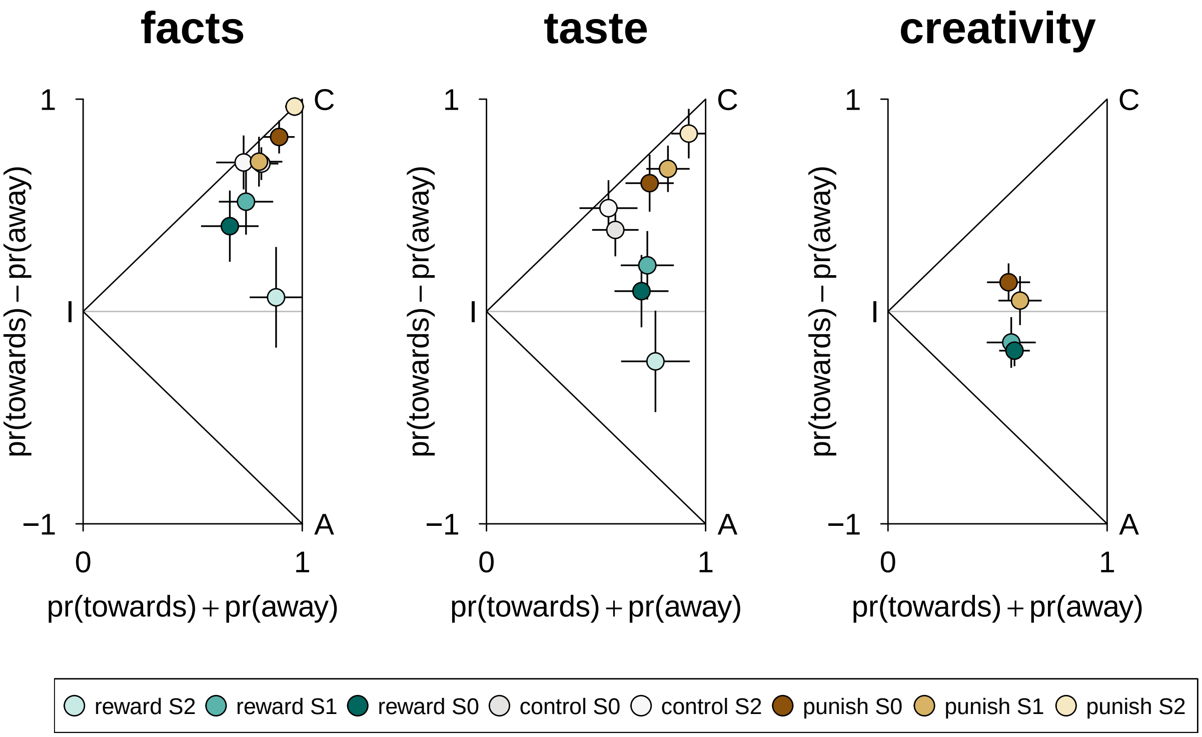


Figure 2: Average response to social influence across treatments

The dots show average behavior in the treatments, whiskers indicate 95% confidence intervals along the two model dimensions based on the t-distribution and block-bootstrapped standard errors, where blocks are subjects in Experiment 1 and matching groups in Experiment 2. Results of the *Creativity* domain rely on minimal Euclidean distance in the RGB color space in the second half of the experiment. (The reason is that by then, participants gained experience with the interface and created a desirable variety of colors allowing them to respond to others' choices. Figure 9 in Appendix B shows that the treatment effects are robust when we use data from all periods as well as alternative measures of similarity.)

The dots on the cyan color scale capture the *Reward* treatments, dots on the orange color scale capture the *Punishment* treatments, and dots on the grey color scale capture the *Control* treatments. The darkest colors refer to our main treatments *S0* with coordination incentives being absent, while brighter colors refer to the *S1* and *S2* treatments fostering the salience rule in the evaluators' decisions.[21]

[21] Throughout our figures, we report results using bootstrapped standard errors. The reason is that in order to estimate standard errors, we need to control for the potential statistical dependence of the choices made by the same participant in Experiment 1 and the choices made in the same matching group in Experiment 2. We report confidence intervals based on block bootstrapped standard errors because

Figure 2 conveys four results. First, and unsurprisingly, we observe ample conformity, with most of the dots in the upper halves of the triangles. Anticonformity is rare, but exists in certain environments (as indicated by the dots in the lower halves of the triangles). Second, comparing the points on the orange and cyan scales shows that the experimental data are consistent with our theoretical predictions. The prospect of *Punishment* creates additional incentives for conformity, whereas the prospect of *Reward* reduces incentives for conformity.

Third, as intended by the design of our domains, conformity tends to be stronger in the *Facts* than in the *Taste* domain, and is weakest in the *Creativity* domain, as shown by the dots moving away from the conformity vertex of the triangle from the left over the middle to the right panel. Fourth, given the similarity of the *S0* and *S1* treatments, coordination incentives to promote salience appear to have little effect. Participants only respond to this rule potentially driving the evaluators' decisions when it is made highly salient as in the *S2* treatments.

Heterogeneity in the response to social influence

The average behavior shown in Figure 2 masks individual differences that may exist in some settings. Figure 3 shows individual differences in responses to incentives, where each dot indicates the response of a behavioral type in a given treatment. The number of behavioral types, their positions, and their frequencies in each treatment are estimated based on mixture models, using the R package *stratEst* (Dvorak, 2023). We use the Bayesian Information Criterion (Schwarz, 1978) to select the number of types. The estimated frequency of each type is represented by the size of its dot. If no dots are shown for a treatment, there are no individual differences, and the behavior is adequately summarized by the average shown in Figure 2.

Figure 3 reveals that heterogeneity exists in some, but not all settings. For the six *Reward* treatments of Experiment 1 (*Reward S0*, *S1*, *S2* in the *Facts* & *Taste* domains), there are two behavioral types in each treatment that differ in their position on the conformity-anticonformity dimension. Most notably, we find an anticonformist type in each *Reward* treatment, with estimated frequencies ranging from 15% to up to 51%. As expected, anticonformity is most frequent and strongest in the *S2* treatments. The more prevalent types are generally conformist, with estimated frequencies ranging from 49% to 85%, and conformity is stronger in the *Facts* domain than in the *Taste* domain.

bootstrapping is straightforward to apply to more complex estimators such as those derived from mixture models. An alternative, statistically equivalent approach are cluster-robust standard errors, which yields very similar results and does not affect our interpretations.

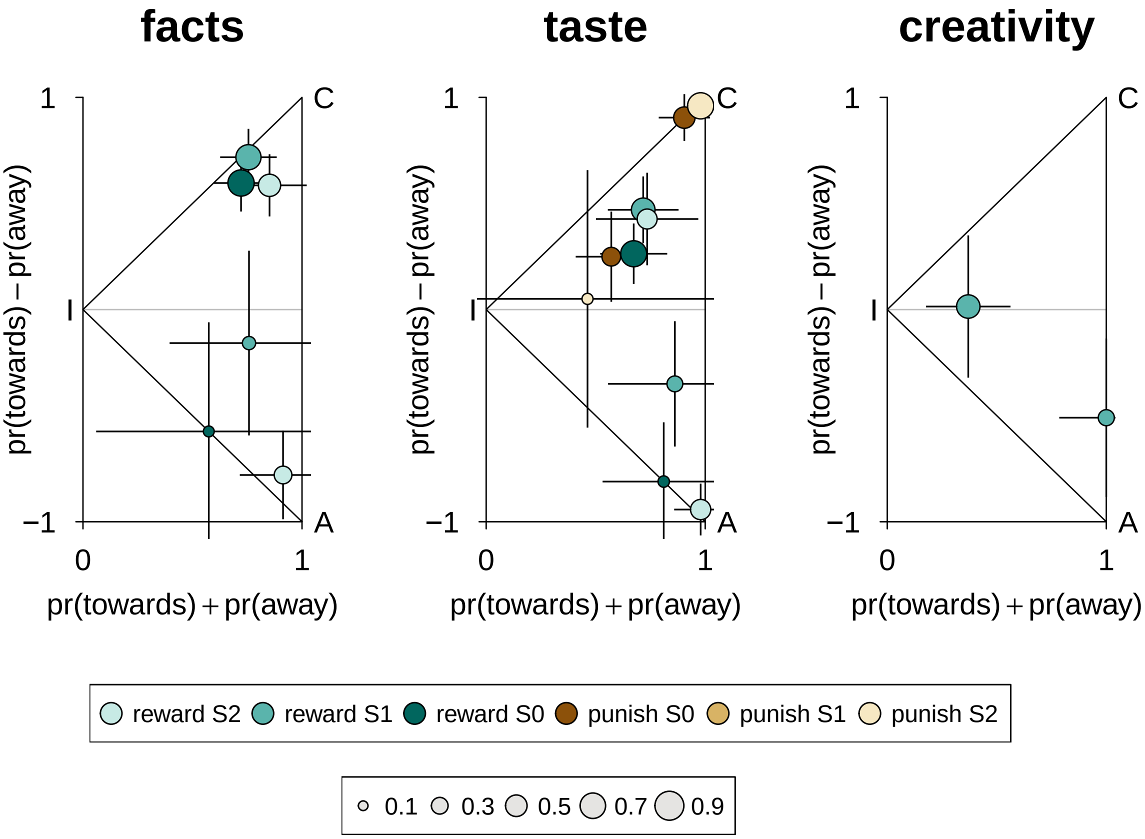


Figure 3: Heterogeneity in types of responses to social influence

The dots indicate the average behavior of the types, and their size captures the estimated frequency of the type in the sample. Whiskers indicate the 95% confidence interval for each type based on the t-distribution and block-bootstrapped standard errors. Number of behavioral types for each treatment that minimizes the Bayesian Information Criterion (Schwarz, 1978).

In the *Creativity* domain, individual differences exists only in the *S1 Reward* treatment, with the most frequent type being independent (68%) and a minority type exhibiting substantial anticonformity. In line with our expectations, the frequency of conformist responses to *Reward* decreases across our domains from *Facts* over *Taste* to *Creativity.*

For the *Punishment* treatments, we observe heterogeneity only in the *Taste* domain in the *S0* and *S2* treatments. In these treatments we find a highly conformist majority type (57% to 85%) and a less conformist minority type (15% to 43%), while anticonformity is absent.

Intransitivity in the informed choices of Experiment 1

Figure 4 shows the relative frequency of intransitivity, i.e., deviations from the predicted choice when being informed about the other group members' choices in Experiment 1 – for example, selecting **Y** instead of the predicted choice **X**. There are three scenarios with respect to the number of other group members who chose in line with a participant's predicted choice.

The scenario where one other group member chose in line and one chose against a participants' predicted choice **X** provides the baseline, as shown by the white dots (**Y**|**X**,XY). The share of baseline intransitivity amounts to about 27 percent (ranging from 20% to 34%) and does not differ systematically across treatments. Accordingly, the concern that people may strategically misreport their actually preferred option in the uninformed decision under *Reward* does not seem to be very relevant.

Participants could adjust their informed choice towards the majority if both other group members selected Y (grey dots, **Y**|**X**,YY), and away from the majority if no other group member selected Y (red dots, **Y**|**X**,XX). To interpret how much conformity and anticonformity a treatment evokes, the baseline in Figure 4 is essential. A treatment induces conformity if the grey dots exceed the frequency of baseline intransitivity, and in the same vein, anticonformity is induced only if the red dots *exceed* baseline intransitivity. Thus, our control treatments show a high level of conformity in the two domains - in terms of increased conformity where conformity is possible (grey dots are above the white dots) and in terms of reduced anticonformity where anticonformity was possible (red dots are below the white dots).

In both domains, the prospect of *Punishment* mainly induces conformity, especially in the *S2* treatments, while it has little effect on anticonformity. In contrast, the prospect of a *Reward* reduces conformity, and increases anticonformity in the *S2* treatment of the *Taste* domain. Thus, our data suggest that while the prevalence of conformity is always affected by the *Reward* and *Punishment* incentives, anticonformity requires specific settings to occur.

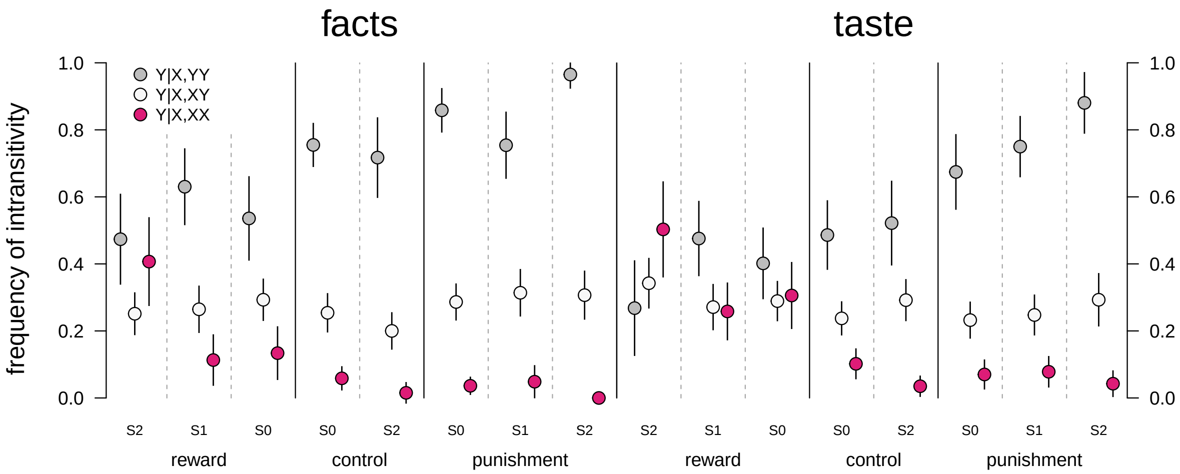


Figure 4: Intransitivity in the informed choices in Experiment 1

Dots show the relative frequency of intransitivity, conditional on the number of other group members with a choice different from the predicted choice. Whiskers indicate 95% CIs of the estimates, based on block-bootstrapped standard errors (10,000 samples, matching subject ID). Note that within each domain, the treatments are ordered according to our hypotheses with respect to the degree of conformity (increasing) and anticonformity (decreasing).

Similarity in the informed choices of Experiment 2

In the *Creativity* domain, we investigate conformity and anticonformity by measuring the degree of similarity between a designer's informed choice (in relation to the four possible options at hand) and the submitted colors of the other three group members. Each designer's shortlist of four colors yields four potential degrees of similarity to the colors of the other group members.

To measure the degree of similarity, we calculate the distance of a designer's color to the colors of the other three group members. We find a higher degree of similarity of the adjusted colors to the colors of the other three group members in the *Punishment* compared to the *Reward* treatments. Table 12 in Appendix B shows that the average distance of the adjusted colors in the *Punishment* treatments is consistently smaller for all our distance measures (see Section B.1 for details on the six different distance measures we use).[22]

To shed light on the mechanisms behind these treatment differences, we identify adjustment strategies by calculating the rank of similarity for each color in a designer's short list to

[22] Figure 10 in Appendix B shows that the average distance decreases over time as participants gain experience over the course of the experiment in the *Punishment* treatments reflecting increasing conformity, and the distance remains rather stable in the *Reward* treatments (except for a drop in the last round).

the submitted colors of the other group members. The highest rank refers to the color with the largest minimum Euclidean distance to the colors of the other three group members (i.e., lowest-rank colors are most similar and highest-rank colors are least similar to the others' colors).

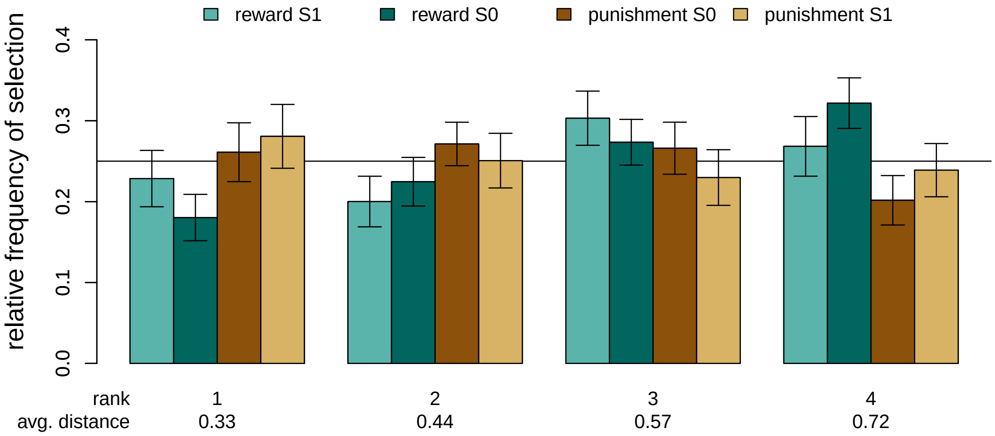


Figure 5: Adjustment to distance ranks in Experiment 2

Relative frequency of distance ranks. The higher the ranks, the larger is the minimum Euclidean distance of the selected color to the others' colors. Whiskers indicate 95% confidence intervals, based on the t-distribution and block-bootstrapped standard errors (10,000 samples, matching group ID).

Figure 5 reveals that the treatment differences are driven by different adjustment strategies. In the *Reward* treatments, designers most frequently adjust to the color with the higher distance ranks 3 and 4 (adjusting away from the others), whereas in the *Punishment* treatments, participants most frequently adjust to the colors of the lower ranks 1 and 2 (adjusting towards the others). The salience-inducing coordination treatments *S1* do not affect these adjustments in a consistent way compared to the main *S0* treatments. The average ranks are 2.74 in the *Reward S0* treatment and 2.41 in the *Punishment S0* treatment (bootstrapped p-value < 0.001). In the *S1* treatments, the average ranks are 2.61 under *Reward* and 2.43 under *Punishment* (bootstrapped p-value $= 0.019$).

Determinants of the adjustment of informed choices

What determines participant's deviation from their predicted choice when being informed about their group members' choices? Some answers are provided in Tables 5 and 6, showing the results of multinomial logit models for the two experiments where a dummy variable indicating an adjustment of the informed choice is regressed on characteristics of the predicted (Experiment 1) or initial (Experiment 2) choice.

In Table 5, the regressor *preference strength* is a continuous variable capturing the predicted strength of the preference for the predicted choice in Experiment 1, measured by the average of the signed preference strengths of the two uninformed choices.[23] The variables *majority choice* and *unique choice*, respectively, are dummies indicating whether the predicted uninformed choice is made by both other group members (**X**,XX) or by no other group member (**X**,YY), respectively. The baseline category refers to the situation where the choices of the other group members differ (**X**,XY).

Table 5: Logit models for the adjustment of choices in Experiment 1

	facts								taste							
	reward			control		punish			reward			control		punish		
	S0	S1	S2	S0	S2	S0	S1	S2	S0	S1	S2	S0	S2	S0	S1	S2
preference	-1.45	-0.75	-1.18	-1.35	-1.32	-1.68	-1.45	-1.06	-1.41	-1.18	-0.99	-2.69	-2.49	-1.20	-1.44	-1.33
strength	(0.62)	(0.58)	(0.47)	(0.39)	(0.80)	(0.64)	(0.58)	(1638.22)	(0.32)	(0.42)	(0.46)	(0.56)	(1.09)	(0.69)	(0.43)	(0.55)
majority	-1.41	-1.79	-0.13	-2.41	-3.85	-2.77	-2.52	-21.23	-0.24	-0.50	0.43	-1.22	-2.59	-2.18	-1.88	-2.65
choice	(0.36)	(0.43)	(0.28)	(0.39)	(7.99)	(0.56)	(1.37)	(2653.01)	(0.27)	(0.29)	(0.31)	(0.29)	(1.43)	(0.39)	(0.42)	(1.34)
unique	0.56	0.82	0.14	1.54	1.28	2.32	1.68	3.65	0.06	0.33	-0.70	0.87	0.93	1.17	1.71	2.56
choice	(0.30)	(0.33)	(0.28)	(0.24)	(0.40)	(0.35)	(0.34)	(11.97)	(0.25)	(0.27)	(0.37)	(0.27)	(0.39)	(0.29)	(0.34)	(0.64)
Obs	284	287	259	356	230	320	287	227	284	285	266	360	235	286	304	233
N	54	51	45	60	42	54	51	39	54	51	45	60	42	54	51	39

Shown are multinomial logit coefficients and block-bootstrapped standard errors in parentheses. The dependent variable is a dummy for intransitivity that takes the value of 1 when the informed choice is adjusted. The independent variables all refer to the predicted choice for Stage 2, based on Stage 1. *Obs* and *N* indicate the number of observations and participants in the sample. Note that the huge standard errors for the *Facts Punishment S2* treatment arise from the fact that heterogeneity is basically absent due to ample conformity in this treatment, as evident from Figure 2.

The coefficients reveal two determinants affecting the probability of intransitivity in the informed choice. First, the stronger a group member's predicted *preference strength* in the predicted informed choice, the less likely intransitivity occurs (as indicated by the consistently negative coefficients across all treatments). *Preference strength* has a statistically significant impact in most of the cases as can be derived from the precisely estimated coefficients. An outlier is the *Facts Punishment S2* treatment where the majority choice is always selected (as evident from Figure 2), irrespective of their *preference strength*.

The second determinant refers to how common the predicted item is in the context of the group, as captured by the variables *majority choice* and *unique choice*. Overall, intransitivity is less likely in cases where the predicted item matches the choices of the other group members

[23] If σ_{XY} is the strength of the preference in favor of X when the alternative is Y then we predict σ_{YZ} as the average of σ_{YX} and $\sigma_{XZ} = -\sigma_{ZX}$.

Table 6: Logit models for the adjustment of choices in Experiment 2

	creativity			
	reward		punish	
	S0	S1	S0	S1
intercept	0.74	0.68	0.27	0.54
	(0.24)	(0.48)	(0.30)	(0.37)
min distance	-0.96	-0.28	0.08	-0.03
	(0.34)	(0.38)	(0.31)	(0.35)
beautiful color	-0.23	-0.24	0.26	-0.29
	(0.25)	(0.44)	(0.38)	(0.44)
interesting color	0.19	-0.14	-0.01	0.09
	(0.44)	(0.35)	(0.44)	(0.41)
Obs	736	640	768	640
N	92	80	96	80

Shown are multinomial logit coefficients and block-bootstrapped standard errors in parentheses. The dependent variable is a dummy that takes the value of 1 when the informed choice is adjusted. The independent variables refer to a designer's initial uninformed choice in Stage 1. The variables reflecting how *beautiful* and *interesting* a designer perceives a color are continuous. *Obs* and *N* indicate the number of observations and participants in the sample.

(indicated by the mostly negative coefficients of *majority choice*), and intransitivity is more likely when the predicted choice stands out (indicated by the mostly positive coefficients of *unique choice*). These coefficients of both variables reflect the prevalence of conformity, and their size varies systematically across treatments in line with the predictions of the theoretical model. Deviating from the predicted majority choice is much *less* likely in the prospect of *Punishment* than in the prospect of *Reward*, whereas deviating from the predicted *unique choice* is much *more* likely in the *Punishment* than in the *Reward* treatments.

These different responses to negative and positive consequences of being selected are particularly pronounced in the *S2* treatments. Deviations from the initial majority choice hardly exist in the *Facts* domain under *Punishment.* In the *Taste* domain under *Reward*, the coefficients of *majority choice* as well as *unique choice* even reverse their signs, implying that intransitivity is more likely if the predicted item is selected by the other group members, and less likely if it has not been selected by others.

Table 6 shows the results of the same exercise for Experiment 2. The variable *min distance* reflects the minimum of the three Euclidean distances of the in Stage 1 initially chosen color to each of the colors chosen by the group members in the RGB color space. The variables capturing how *beautiful* and *interesting* a color is refer to a designer's perception of their uninformed color choice and rest on their ratings on continuous scales ranging from zero (not beautiful at all / not interesting at all) to one (very beautiful / very interesting).

In the *Creativity* domain, the logit coefficients indicate that the decision to adjust the choice under social influence is affected by strategic considerations in the *Reward*, but not in the *Punishment* treatments. Participants more frequently adjust their informed choice in the *Reward* treatment the more similar their initially chosen color is to the color of a group member, as captured by the negative coefficients of the minimum Euclidean distance. This effect is substantial in the *S0* treatment and weak but qualitatively in the same direction in the *S1* treatment. Post-experimental ratings of the designers regarding how beautiful and interesting they perceive their created colors do not predict the decision to adjust the informed choice. Thus, in the *Creativity* domain, designers attempt to stand out in the prospect of a *Reward*. They neither try to hide in the majority in the prospect of *Punishment*, nor do they act according to homophily-driven evaluation.

5.2 Evaluators' decisions: homophily and salience

As outlined in our model, homophily and salience may both drive the evaluators' decisions. Disentangling the relative importance of the two is essential to understand the incentives for conformity and anticonformity in our various treatments. Figures 6 and 7 give an impression of the importance of the two mechanisms. Note that the percentages of the two figures may add up to more than 1 because the item matching the evaluator's own choice (homophily) may coincide with the single item (salience). We complement the figures with regression analyses.

Evaluators' frequencies of rewarding and punishing based on homophily

Figure 6 shows evaluators' selection decisions based on homophily across the treatments including evaluation. The left and central panels show the relative frequencies of selecting the answer and painting that matches the evaluators' choice when they themselves decided in the role of a group member. The panel on the right shows the relative frequency of evaluators selecting the color they rated as being most beautiful.

The figure shows that homophily is a very powerful mechanism driving the evaluators' selections. In the *Facts* and *Taste* domains, evaluators select the item matching their own prior choice for a *Reward* in about 80% of the cases in the *S0* and *S1* treatments, and even in about 60% in the *S2* treatments. Homophily is also the predominant mechanism for assigning a *Reward* in the *Creativity* domain.

When assigning *Punishment* in the absence of coordination incentives (*S0*), evaluators avoid the item they like themselves in all domains. This mechanism is mitigated in the *S1*

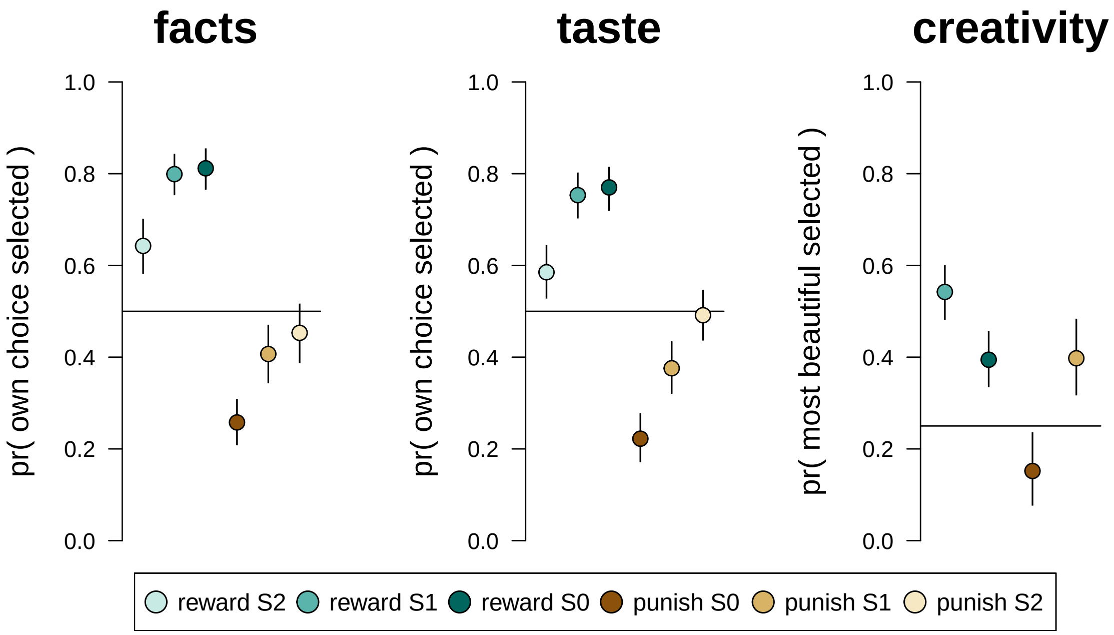


Figure 6: Selection probabilities of the item matching the evaluator's own preference

In the left and central panel, the dots indicate the probability of the evaluator selecting the answer or question which matches the evaluators' own prior choice in the role of a group member. In the panel on the right, the dots indicate an evaluators' probability of selecting the color she/he rated as most beautiful. Whiskers indicate block-bootstrapped 95% confidence intervals based on the t-distribution and block-bootstrapped standard errors. Horizontal lines mark the expected probability in case of random selections.

and *S2* treatments.

Evaluators' frequencies of rewarding and punishing based on salience

Figure 7 shows evaluators' selection decisions based on salience. The left and central panel show the relative frequencies of selecting the answer and painting that stands out, i.e., the item chosen by only one group member. The panel on the right shows the relative frequency evaluators selecting the color with the largest minimal distance to the colors of the other group members.

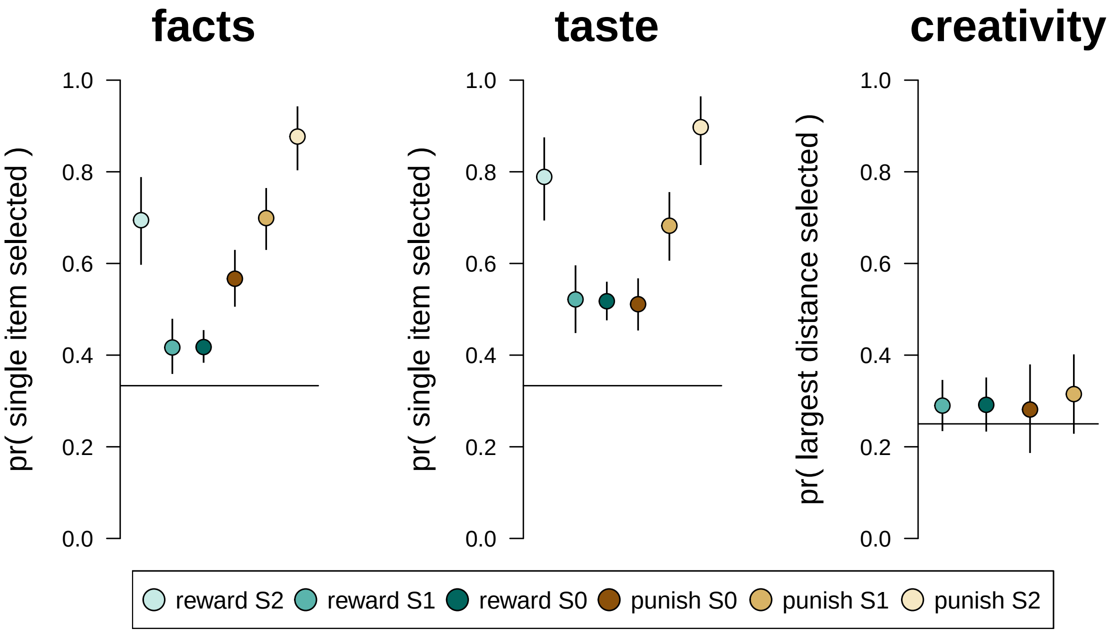


Figure 7: Selection probabilities of the item standing out

In the left and central panel, the dots indicate the probability of the evaluator selecting the single answer or question. In the panel on the right, the dots indicate the evaluator's probability of selecting the color with the largest minimal distance. Whiskers indicate block-bootstrapped 95% confidence intervals based on the t-distribution and block-bootstrapped standard errors. Horizontal lines mark the expected probability in case of random selections.

In the *Facts* and *Taste* domains, the frequency of the single item being selected is always larger than the one-third frequency in case of random allocations, implying that standing out always increases the chance to attract a reward, and punishment can be avoided by "hiding" in the majority.

In the absence of coordination incentives (*S0*), in the *Facts* domain, evaluators select the single answer more frequently when punishing instead of rewarding, whereas in the *Taste* domain, the positive or negative consequences of their selection do not affect the probability of the single painting being selected. Moreover, while evaluators choose the single painting in slightly more than half of the cases in both *S0* treatments of the *Taste* domain, the single answer to *Facts* questions in case of *Punishment* is selected less frequently by about 10%. These observations are in line with the interpretation that evaluators may perceive the less frequently chosen answer in a group as a negative signal about its correctness, while there is no truth involved in subjective arts preferences.

Comparing the *S0* and *S1* treatments reveals that coordination incentives increase the probability of the single item being selected only in the *Punishment*, but not in the *Reward* treatments of both domains. We will return to this observation at the end of this section. Eventually, the *S2* treatments succeed in triggering the salience rule as evaluators assign *Punishment* to the single item in about 90% of their choices, and they assign *Reward* to the single answer in about 70% and to the single painting in about 80% of their choices.

In the *Creativity* domain (right panel of Figure 7), incentives for standing out are small as the relative frequencies of the evaluator selecting the color with the largest minimal distance are not much larger than the one-fourth expected frequency in case of random selections, and they are unaffected by the treatments. In what follows, we contrast the potential impact of salience with homophily to better understand the evaluators' decisions.

Salience-based versus homophily-based evaluation

We investigate the importance of salience and homophily for the evaluators' decisions relying on conditional logit models (McFadden, 1974), where the choice among several alternatives is modeled as a function of the characteristics of the alternatives. To analyze the extent to which evaluators allocate *Reward* and *Punishment* based on salience or homophily, we regress a dummy variable indicating the evaluator's selected item on characteristics of the evaluated choices. We use the R package *mlogit* (Croissant, 2019) to obtain maximum likelihood estimates for the model coefficients and block-bootstrapped standard errors. The regression results are shown in Table 7.

The models referring to Experiment 1 contain the dummy variables *salience* (indicating the single answer or painting), and *homophily* (indicating whether or not the selected item equals the evaluator's own prior choice in the role of a group member).

The regressors predicting the evaluator's selection in Experiment 2 were elicited after the main experiment. The variable *salient color* is a dummy indicating the color generally considered to be salient, as derived from a Krupka-Weber coordination task (Krupka and Weber, 2013) among all participants across the four colors. The variables capturing how *beautiful* and *interesting* a color is refer the evaluator's perception and rest on their ratings

Table 7: Salience-based versus homophily-based evaluation

	Facts						Taste						Creativity			
	reward			punishment			reward			punishment			reward		punishment	
	S0	S1	S2	S0	S1	S2	S0	S1	S2	S0	S1	S2	S0	S1	S0	S1
salience	-0.50	-0.30	0.90	0.26	0.93	1.93	0.10	0.07	1.39	0.14	0.76	2.32				
	(0.08)	(0.10)	(0.08)	(0.07)	(0.08)	(0.12)	(0.08)	(0.06)	(0.09)	(0.10)	(0.10)	(0.10)				
homophily	1.53	1.40	0.74	-1.07	-0.44	-0.28	1.21	1.12	0.56	-1.26	-0.57	-0.25				
	(0.11)	(0.10)	(0.13)	(0.16)	(0.14)	(0.11)	(0.12)	(0.13)	(0.11)	(0.07)	(0.11)	(0.08)				
salient color													1.11	1.25	0.25	1.17
													(0.19)	(0.15)	(0.23)	(0.17)
beautiful color													1.09	1.91	-2.15	0.35
													(0.56)	(0.42)	(0.58)	(0.40)
interesting color													0.56	1.07	-1.22	0.25
													(0.56)	(0.48)	(0.91)	(0.48)
Obs	648	612	540	648	612	468	648	612	540	648	612	468	736	1,280	768	1,280
N	54	51	45	54	51	39	54	51	45	54	51	39	23	40	24	40
LL	-302	-303	-308	-366	-357	-177	-349	-342	-269	-342	-367	-142	-213	-308	-234	-383

Conditional logit regression with the evaluator's selected item as the dependent variable. Coefficients and block-bootstrapped standard errors are in parentheses. For the *Facts* and *Taste* domains, the variable *salience* is a dummy indicating the single item, and *homophily* is a dummy indicating the evaluator's own choice in the role of a group member. For the *Creativity* domain, the *salient color* is identified by the minimum of the Euclidean distances to the other three colors. The variables reflecting how *beautiful* and *interesting* an evaluator perceives a color are continuous. The labels *Obs*, *N* and *LL* refer to the number of observations, participants, and log likelihood of the model, respectively.

on continuous scales ranging from zero (not beautiful at all / not interesting at all) to one (very beautiful / very interesting).

Table 7 reveals that homophily is a powerful and robust mechanism determining the allocation of *Reward* and *Punishment*. The coefficients of the variables related to homophily are always positive for the *Reward* treatments and largely negative for the *Punishment* treatments. The effect sizes are substantial and precisely estimated, especially across all treatments in Experiment 1, and stronger compared to the coefficients reflecting salience (particularly so in *S0*). An exception is the *Punishment S1* treatment in the *Creativity* domain, where homophily (based on the evaluators' perceptions of how beautiful and interesting a color is) does not escape being selected for punishment, suggesting a minor role of homophily-based allocation in this particular condition involving coordination incentives. Overall, evaluators favor those who appear similar to themselves – they are more likely to receive a reward and less likely to be punished.

In contrast, salience-based allocation of reward and punishment is a much weaker mechanism. If punishment and reward were allocated based on salience, the coefficients of the corresponding variables should be positive. In our main treatments where coordination incentives are absent, the coefficients of the salience variables in Table 7 are generally small, indicating that salience plays a minor role. The coefficients for *salience* are even negative in the *Facts* domain, which is consistent with the interpretation that evaluators may attempt to punish wrong answers (believing that the majority answer is more likely to be correct).

In the *S1* treatments, the coefficients of the variables reflecting salience show that our experimental manipulation to induce salience-based evaluation worked in the *Punishment* treatments of Experiment 1 and Experiment 2 but failed in the *Reward* treatments of both experiments, corroborating our observation from Figure 7.

Two potential explanations may account for these different effects of the coordination incentives in the *Punishment* and *Reward* treatments. First, to allocate *Punishment*, selecting the salient item and not the item matching the evaluator's own choice as a group member is plausible. However, evaluators may still prefer to allocate a *Reward* to someone who decides like themselves (i.e., based on homophily) – regardless of the coordination incentives meant to foster salience.

Second, while the *S1* treatments create incentives for evaluators to coordinate, salience in terms of the single item may not be the only coordination criterion. For example, in the *Punishment* treatment of the *Facts* domain, evaluators could also attempt to coordinate on the assumingly wrong answer. Even though we constructed the answer options such that the correct answers are hardly known by anyone, evaluators may perceive the frequency of a selected answer as a signal about its correctness. Thus, coordinating on the single item coincides with coordinating on the probably wrong answer and thus increases the probability of the single item being selected from *S0* to *S1*. In the *Reward* treatment of the *Facts* domain, the evaluation criterion is ambiguous: evaluators may coordinate on the single (but maybe wrong) answer, or on the probably correct (but majority) answer, which may explain that the *S1* treatment does not differ from *S0* in this case.

Adding the mechanistic element of training participants to coordinate on the single item in *S2* finally gets salience to work and diminishes the relative importance of homophily for negative as well as positive consequences of being selected, and in the two domains of Experiment 1.

In a nutshell, evaluators' selection decisions suggest that homophily is a very powerful driving force, whereas salience needs to be very salient in order to take effect. Accordingly, matching the evaluator's taste may be more important than standing out to avoid punishment and attract a reward, which is in line with the group members' choices described in Subsection 5.1.

6 Discussion

The effect of incentives on the balance between conformist and anticonformist behavior has received little attention in the literature on social influence. In this paper, we theoretically and experimentally show that evaluation can not only incentivize conformity, but also anticonformity. In a theoretical model we analyse how punishment and reward affect conformity and anticonformity. For both decision rules of the evaluator that we investigate - salience and homophily, we show that punishment creates incentives for conformity. Reward compromises conformity and can even create incentives for anticonformity. In two laboratory experiments, we find that evaluation induces strategic conformity in the case of punishment, and strategic anticonformity in the case of reward. The effects of evaluation are consistent across three choice domains, despite varying levels of baseline conformity across the domains.

Our experiments also shed light on the mechanisms driving evaluators' selection decisions. We find that homophily is a much more powerful rule for assigning reward and punishment than salience (unless the latter is pushed to an extreme by design). This finding calls for an extension of Rubinstein's quote, recommending to deviate from the common dress code for the reason of homophily in addition to salience. As typically, the applicant is uncertain about the evaluator's preferences over clothing styles, our study would imply that: *In the likely event that the evaluator prefers the conventional look, the casual outfit does not significantly hurt your chances to get the job because they are small anyway. However, in the unlikely event that you meet an unconventional evaluator like Rubinstein, the casual look will clearly* *increase your chances.*

The individual level aside, societal benefits and costs of conformity and anticonformity may vary considerably across situations, determining the use of positive or negative incentives in a specific context. Conformity can be exploited for desirable economic outcomes, in particular norm compliance.[24] On the downside, conformity may be the reason for why people make irrational financial decisions, shy away from innovative practices, are susceptible to group-think, or communicate in echo chambers or filter bubbles. Instead, anticonformity may lead to new practices and discoveries in organizations as well as societies, break information cascades and erode archaic social conventions - but also reduce coordination and predictability of behavior, which can be detrimental for a society. Thus, the potential of incentives for conformity and anticonformity looms large.

References

Allcott, Hunt. 2011. "Social norms and energy conservation." *Journal of Public Economics* 95 (9):1082 – 1095.

Allcott, Hunt and Todd Rogers. 2014. "The Short-Run and Long-Run Effects of Behavioral Interventions: Experimental Evidence from Energy Conservation." *American Economic Review* 104 (10):3003–37.

Allen, Vernon L. 1965. "Situational Factors In Conformity." *Advances in Experimental Social Psychology* 2:133 – 175. URL `http://www.sciencedirect.com/science/article/pii/S0065260108601057`.

Alpizar, Francisco, Fredrik Carlsson, and Olof Johansson-Stenman. 2008. "Anonymity, reciprocity, and conformity: Evidence from voluntary contributions to a national park in Costa Rica." *Journal of Public Economics* 92 (5):1047 – 1060. URL `http://www.sciencedirect.com/science/article/pii/S0047272707001909`.

Amabile, Teresa M., Phyllis Goldfarb, and Shereen C. Brackfleld. 1990. "Social influences on creativity: Evaluation, coaction, and surveillance." *Creativity Research Journal* 3 (1):6–21. URL `https://doi.org/10.1080/10400419009534330`.

[24] When being informed that a majority of other people will conform, people are more likely to pay taxes (Bobek, Roberts, and Sweeney, 2007; Coleman, 2007), save energy (Allcott, 2011; Allcott and Rogers, 2014; Nolan et al., 2008; Schultz et al., 2007), donate to a charity (Alpizar, Carlsson, and Johansson-Stenman, 2008; Smith, Windmeijer, and Wright, 2015) and contribute to a political party (Perez-Truglia and Cruces, 2017).

Amini, Makan, Mathias Ekström, Tore Ellingsen, Magnus Johannesson, and Fredrik Strömsten. 2017. "Does gender diversity promote nonconformity?" *Management Science* 63 (4):1085–1096.

Anderson, Lisa R. and Charles A. Holt. 1997. "Information Cascades in the Laboratory." *American Economic Review* 87 (5):847–862. URL http://www.jstor.org/stable/2951328.

Apesteguia, Jose, Steffen Huck, and Jörg Oechssler. 2007. "Imitation — theory and experimental evidence." *Journal of Economic Theory* 136 (1):217–235.

Argyle, M. . 1957. "Social pressure in public and private situations." *Journal of Abnormal and Social Psychology* 54:172–175.

Ariel Rubinstein. 2013. "10 Q&A: Experienced Advice forp "Lost" Graduate Students in Economics." *The Journal of Economic Education* 44 (3):193–196. URL https://doi.org/10.1080/00220485.2013.795448.

Ariely, Dan and Jonathan Levav. 2000. "Sequential Choice in Group Settings: Taking the Road Less Traveled and Less Enjoyed." *Journal of Consumer Research* 27 (3):279–290. URL https://doi.org/10.1086/317585.

Asch, S. 1951. *Groups, leadership, and men*, chap. Effects of group pressure upon the modification and distortion of judgments. Pittsburgh, PA: Carnegie Press, 177–190.

Asch, S. E. 1955. "Opinions and social pressure." *Scientific American* 193:31–35.

Asch, S.E. 1952. *Social Psychology*. New-Jersey: Prentice-Hall.

Baccara, Mariagiovanna and Leeat Yariv. 2013. "Homophily in Peer Groups." *American Economic Journal: Microeconomics* 5 (3):69–96.

Banerjee, Abhijit and Drew Fudenberg. 2004. "Word-of-mouth learning." *Games and Economic Behavior* 46 (1):1 – 22. URL http://www.sciencedirect.com/science/article/pii/S0899825603000484.

Banerjee, Abhijit V. 1992. "A Simple Model of Herd Behavior." *The Quarterly Journal of Economics* 107 (3):797–817. URL http://dx.doi.org/10.2307/2118364.

Becker, Gary. 1971. *The Economics of Discrimination*. University of Chicago Press, 2 ed. URL https://EconPapers.repec.org/RePEc:ucp:bkecon:9780226041162.

Belloc, M. and S. Bowles. 2013. "The Persistence of Inferior Cultural-Institutional Conventions." *American Economic Review* 103 (3):93–98.

Bernheim, B. D. 1994. "A Theory of Conformity." *Journal of Political Economy* 102 (5):841–877. URL https://doi.org/10.1086/261957.

Bertrand, M. and E. Duflo. 2017. "Field Experiments on Discrimination." In *Handbook of Field Experiments*, *Handbook of Economic Field Experiments*, vol. 1, edited by Abhijit Vinayak Banerjee and Esther Duflo. North-Holland, 309 – 393. URL http://www.sciencedirect.com/science/article/pii/S2214658X1630006X.

Bikhchandani, Sushil, David Hirshleifer, and Ivo Welch. 1992. "A Theory of Fads, Fashion, Custom, and Cultural Change as Informational Cascades." *Journal of Political Economy* 100 (5):992–1026. URL https://doi.org/10.1086/261849.

Bobek, D., R. Roberts, and J. Sweeney. 2007. "The social norms of tax compliance: evidence from Australia, Singapore and the United States." *Journal of Business Ethics* 74 (1):49–64.

Bock, Olaf, Ingmar Baetge, and Andreas Nicklisch. 2014. "hroot: Hamburg registration and organization online tool." *European Economic Review* 71:117–120.

Boehm, Christopher. 1993. "Egalitarian behavior and reverse dominance hierarchy." *Current Anthropology* 34 (3):227–254. URL <GotoISI>://WOS:A1993LC40800002.

———. 2000. *Hierarchy in the Forest: The Evolution of Egalitarian Behavior*. Cambridge, MA: Harvard University Press.

Bond, Michael Harris and Peter B. Smith. 1996. "Cross-Cultural Social and Organizational Psychology." *Annual Review of Psychology* 47 (1):205–235. URL https://doi.org/10.1146/annurev.psych.47.1.205. PMID: 15012481.

Bordalo, Pedro, Nicola Gennaioli, and Andrei Shleifer. 2022. "Salience." *Annual Review of Economics* 14 (1):521–544.

Bowles, Samuel and Jung-Kyoo Choi. 2013. "Coevolution of farming and private property during the early Holocene." *Proceedings of the National Academy of Sciences* 110 (22):8830–8835. URL https://www.pnas.org/content/pnas/110/22/8830.full.pdf.

Boyd, R. and P. J. Richerson. 1985. *Culture and the evolutionary process*. Chicago: University of Chicago Press.

Boyd, Robert, Herbert Gintis, and Samuel Bowles. 2010. "Coordinated Punishment of Defectors Sustains Cooperation and Can Proliferate When Rare." *Science* 328 (5978):617–620. URL https://science.sciencemag.org/content/sci/328/5978/617.full.pdf.

Boyd, Robert and Peter J. Richerson. 1982. "Cultural transmission and the evolution of cooperative behavior." *Human Ecology* 10 (3):325–351. URL https://doi.org/10.1007/BF01531189.

Brehm, J. W. 1966. *A theory of psychological reactance*. Oxford, England: Academic Press.

Buss, David M. 2003. *The evolution of desire: Strategies of human mating*. New York: Basic Books, 2nd ed.

Byrne, Donn Erwin. 1971. *The Attraction Paradigm*. New York, NY: Academic Press.

Charness, Gary, Michael Naef, and Alessandro Sontuoso. 2019. "Opportunistic conformism." *Journal of Economic Theory* 180:100–134.

Cialdini, Robert B. and Noah J. Goldstein. 2004. "Social Influence: Compliance and Conformity." *Annual Review of Psychology* 55 (1):591–621. URL https://doi.org/10.1146/annurev.psych.55.090902.142015. PMID: 14744228.

Cialdini, Robert B., Wilhelmina Wosinska, Daniel W. Barrett, Jonathan Butner, and Malgorzata Gornik-Durose. 1999. "Compliance with a Request in Two Cultures: The Differential Influence of Social Proof and Commitment/Consistency on Collectivists and Individualists." *Personality and Social Psychology Bulletin* 25 (10):1242–1253. URL https://doi.org/10.1177/0146167299258006.

Clark, Philip J. and Francis C. Evans. 1954. "Distance to Nearest Neighbor as a Measure of Spatial Relationships in Populations." *Ecology* 35 (4):445–453. URL https://esajournals.onlinelibrary.wiley.com/doi/abs/10.2307/1931034.

Coleman, Stephen. 2007. "The Minnesota Income Tax Compliance Experiment: Replication of the Social Norms Experiment." Tech. rep., Available at SSRN: https://ssrn.com/abstract=1393292 or http://dx.doi.org/10.2139/ssrn.1393292.

Corazzini, Luca and Ben Greiner. 2007. "Herding, social preferences and (non-)conformity." *Economics Letters* 97 (1):74 – 80. URL http://www.sciencedirect.com/science/article/pii/S0165176507000602.

Croissant, Yves. 2019. *mlogit: Multinomial Logit Models*. URL https://CRAN.R-project.org/package=mlogit. R package version 0.4-1.

Crutchfield, R. S. 1962. *Contemporary approaches to creative thinking*, chap. Conformity and creative thinking. New York, NY: Atherton, 120–140.

Crutchfield, Richard S. 1955. "Conformity and character." *American Psychologist* 10 (5):191.

Currarini, Sergio, Matthew O Jackson, and Paolo Pin. 2009. "An economic model of friendship: Homophily, minorities, and segregation." *Econometrica* 77 (4):1003–1045.

Denton, Kaleda Krebs, Yoav Ram, Uri Liberman, and Marcus W. Feldman. 2020. "Cultural evolution of conformity and anticonformity." *Proceedings of the National Academy of Sciences* 117 (24):13603–13614.

Dvorak, Fabian. 2023. "stratEst: a software package for strategy frequency estimation." *Journal of the Economic Science Association* 9 (2):337–349. URL https://doi.org/10.1007/s40881-023-00141-7.

Efferson, Charles, Rafael Lalive, Peter J. Richerson, Richard McElreath, and Mark Lubell. 2008. "Conformists and mavericks: the empirics of frequency-dependent cultural transmission." *Evolution and Human Behavior* 29 (1):56–64. URL http://dx.doi.org/10.1016/j.evolhumbehav.2007.08.003.

Ertug, Gokhan, Julia Brennecke, Balázs Kovács, and Tengjian Zou. 2022. "What Does Homophily Do? A Review of the Consequences of Homophily." *Academy of Management Annals* 16 (1):38–69.

Falk, Armin and Florian Zimmermann. 2017. "Consistency as a Signal of Skills." *Management Science* 63 (7):2197–2210. URL https://doi.org/10.1287/mnsc.2016.2459.

Fatas, Enrique, Shaun P. Hargreaves Heap, and David Rojo Arjona. 2018. "Preference conformism: An experiment." *European Economic Review* 105:71 – 82. URL http://www.sciencedirect.com/science/article/pii/S0014292118300412.

Fehr, Ernst and Urs Fischbacher. 2004. "Social norms and human cooperation." *Trends in Cognitive Sciences* 8 (4):185 – 190. URL http://www.sciencedirect.com/science/article/pii/S1364661304000506.

Fehr, Ernst, Urs Fischbacher, and Simon Gächter. 2002. "Strong reciprocity, human cooperation, and the enforcement of social norms." *Human Nature* 13 (1):1–25. URL https://doi.org/10.1007/s12110-002-1012-7.

Festinger, L. 1953. *Group relations at the crossroads*, chap. An analysis of compliant behavior. Harper, 232–256.

Fischbacher, Urs. 2007. "Z-Tree: Zurich toolbox for ready-made economic experiments." *Experimental Economics* 10 (2):171–178.

Fromkin, H. L. 1970. "Effects of experimentally aroused feelings of undistinctiveness upon valuation of scarce and novel experiences." *Journal of Personality and Social Psychology* 16 (3):521–529. URL `https://doi.org/10.1037/h0030059`.

Gama, Jose and Glenn Davis. 2018. *colorscience: Color Science Methods and Data.* URL `https://CRAN.R-project.org/package=colorscience`. R package version 1.0.5.

Goeree, Jacob K. and Leeat Yariv. 2015. "Conformity in the lab." *Journal of the Economic Science Association* 1 (1):15–28. URL `https://doi.org/10.1007/s40881-015-0001-7`.

Goldberg, A. and S. K. Stein. 2018. "Beyond Social Contagion: Associative Diffusion and the Emergence of Cultural Variation." *American Sociological Review* 83 (5):897–932.

Golub, Benjamin and Matthew O. Jackson. 2012. "How Homophily Affects the Speed of Learning and Best-Response Dynamics." *The Quarterly Journal of Economics* 127 (3):1287–1338. URL `https://doi.org/10.1093/qje/qjs021`.

Goncalo, J. A. and B. M. Staw. 2006. "Individualism-collectivism and group creativity." *Organizational Behavior and Human Decision Processes* 100 (1):96–109.

Greiner, Ben. 2015. "Subject pool recruitment procedures: organizing experiments with ORSEE." *Journal of the Economic Science Association* 1 (1):114–125.

Griskevicius, Vladas, Noah J. Goldstein, Chad R. Mortensen, Robert B. Cialdini, and Douglas T. Kenrick. 2006. "Going Along Versus Going Alone: When Fundamental Motives Facilitate Strategic (Non)Conformity." *Journal of Personality and Social Psychology* 91 (2):281–294.

Guarino, A., H. Harmgart, and S. Huck. 2011. "Aggregate information cascades." *Games and Economic Behavior* 73 (1):167–185. URL `<GotoISI>://WOS:000294579800011`.

Gürerk, Özgür, Bernd Irlenbusch, and Bettina Rockenbach. 2006. "The Competitive Advantage of Sanctioning Institutions." *Science* 312 (5770):108–111.

Guzmán, Ricardo Andrés, Carlos Rodriguez-Sickert, and Robert Rowthorn. 2007. "When in Rome, do as the Romans do: the coevolution of altruistic punishment, conformist learning, and cooperation." *Evolution and Human Behavior* 28 (2):112 – 117. URL http://www.sciencedirect.com/science/article/pii/S1090513806000663.

Hamilton, W.D. 1971. "Geometry for the selfish herd." *Journal of Theoretical Biology* 31 (2):295–311. URL https://www.sciencedirect.com/science/article/pii/0022519371901895.

Hegde, D. and J. Tumlinson. 2014. "Does Social Proximity Enhance Business Partnerships? Theory and Evidence from Ethnicity's Role in US Venture Capital." *Management Science* 60 (9):2355–2380.

Henrich, Joseph, Richard McElreath, Abigail Barr, Jean Ensminger, Clark Barrett, Alexander Bolyanatz, Juan Camilo Cardenas, Michael Gurven, Edwins Gwako, Natalie Henrich, Carolyn Lesorogol, Frank Marlowe, David Tracer, and John Ziker. 2006. "Costly Punishment Across Human Societies." *Science* 312 (5781):1767–1770. URL https://science.sciencemag.org/content/312/5781/1767.

Herrmann, Benedikt, Christian Thöni, and Simon Gächter. 2008. "Antisocial Punishment Across Societies." *Science* 319 (5868):1362–1367. URL http://www.sciencemag.org/cgi/content/abstract/319/5868/1362.

Hertwig, Ralph and Andreas Ortmann. 2001. "Experimental practices in economics: A methodological challenge for psychologists?" *Behavioral and Brain Sciences* 24 (3):383–451. URL <GotoISI>://000171155500001.

Hewstone, Miles, Mark Rubin, and Hazel Willis. 2002. "Intergroup Bias." *Annual Review of Psychology* 53 (1):575–604. URL https://doi.org/10.1146/annurev.psych.53.100901.135109. PMID: 11752497.

Imhoff, Roland and Hans-Peter Erb. 2009. "What Motivates Nonconformity? Uniqueness Seeking Blocks Majority Influence." *Personality and Social Psychology Bulletin* 35 (3):309–320. URL https://doi.org/10.1177/0146167208328166. PMID: 19098256.

Jones, Daniel and Sera Linardi. 2014. "Wallflowers: Experimental Evidence of an Aversion to Standing Out." *Management Science* 60 (7):1757–1771. URL https://doi.org/10.1287/mnsc.2013.1837.

Jones, Stephen R. G. 1984. *The Economics of Conformism.* Oxford: Blackwell.

Juul, Jonas S. and Mason A. Porter. 2019. “Hipsters on networks: How a minority group of individuals can lead to an antiestablishment majority.” *Physical Review E* 99 (2). URL `http://dx.doi.org/10.1103/PhysRevE.99.022313`.

Kelman, Herbert C. 1961. “Processes of Opinion Change.” *Public Opinion Quarterly* 25:57–78.

Kets, Willemien and Alvaro Sandroni. 2016. “Challenging Conformity: A Case for Diversity.” Tech. rep., SSRN Working Paper. Available at SSRN: https://ssrn.com/abstract=2871490 or http://dx.doi.org/10.2139/ssrn.2871490.

———. 2021. “A Theory of Strategic Uncertainty and Cultural Diversity.” *The Review of Economic Studies* 88 (1):287–333.

Kim, Heejung and Hazel Rose Markus. 1999. “Deviance or uniqueness, harmony or conformity? A cultural analysis.” *Journal of Personality and Social Psychology* 77 (4):785–800.

Krupka, Erin L. and Roberto A. Weber. 2013. “Identifying Social Norms Using Coordination Games: Why Does Dictator Game Sharing Vary?” *Journal of the European Economic Association* 11 (3):495–524. URL `https://doi.org/10.1111/jeea.12006`.

Lazarsfeld, P. F. and R. K. Merton. 1954. *Friendship as a Social Process: A Substantive and Methodological Analysis*. New York: Van Nostrand, 18–66.

Lee, In Ho. 1993. “On the Convergence of Informational Cascades.” *Journal of Economic Theory* 61 (2):395 – 411. URL `http://www.sciencedirect.com/science/article/pii/S0022053183710744`.

———. 1998. “Market Crashes and Informational Avalanches.” *The Review of Economic Studies* 65 (4):741–759. URL `https://doi.org/10.1111/1467-937X.00066`.

Loomes, Graham, Chris Starmer, and Robert Sugden. 1991. “Observing violations of transitivity by experimental methods.” *Econometrica: Journal of the Econometric Society* :425–439.

Lynn, Michael and Judy Harris. 1997. “Individual Differences in the Pursuit of Self-Uniqueness Through Consumption.” *Journal of Applied Social Psychology* 27 (21):1861–1883. URL `https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1559-1816.1997.tb01629.x`.

Mahdi, N. Q. 1986. “Pukhtunwali: Ostracism and honor among the Pathan Hill tribes.” *Ethology and Sociobiology* 7 (3-4):295–304. URL `<GotoISI>://WOS:A1986E338800014`.

Mäkelä, K., I. Björkman, and M. Ehrnrooth. 2010. “How do MNCs establish their talent pools? Influences on individuals’ likelihood of being labeled as talent.” *Journal of World Business* 45 (2):134–142.

March, James G. 1991. “Exploration and Exploitation in Organizational Learning.” *Organization Science* 2 (1):71–87. URL `https://doi.org/10.1287/orsc.2.1.71`.

Matusik, S. F., J. M. George, and M. B. Heeley. 2008. “VALUES AND JUDGMENT UNDER UNCERTAINTY: EVIDENCE FROM VENTURE CAPITALIST ASSESSMENTS OF FOUNDERS.” *Strategic Entrepreneurship Journal* 2 (2):95–115.

McElreath, Richard, Adrian V. Bell, Charles Efferson, Mark Lubell, Peter J. Richerson, and Timothy Waring. 2008. “Beyond existence and aiming outside the laboratory: estimating frequency-dependent and pay-off-biased social learning strategies.” *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* 363 (1509):3515–3528. URL `https://www.ncbi.nlm.nih.gov/pubmed/18799416`.

McFadden, D. 1974. *Frontiers in Econometrics*, chap. Conditional logit analysis of qualitative choice behavior. Academic Press, New York, 105–142.

McPherson, Miller, Lynn Smith-Lovin, and James M Cook. 2001. “Birds of a Feather: Homophily in Social Networks.” *Annual Review of Sociology* 27 (1):415–444.

Mehta, Judith, Chris Starmer, and Robert Sugden. 1994a. “Focal points in pure coordination games: An experimental investigation.” *Theory and Decision* 36 (2):163–185. URL `https://doi.org/10.1007/BF01079211`.

———. 1994b. “The Nature of Salience: An Experimental Investigation of Pure Coordination Games.” *The American Economic Review* 84 (3):658–673. URL `http://www.jstor.org/stable/2118074`.

Nail, P. R. 1986. “Toward an integration of some models and theories of social response.” *Psychological Bulletin* 100 . (2):190–206.

Nail, Paul R., Stefano I. Di Domenico, and Geoff MacDonald. 2013. “Proposal of a Double Diamond Model of Social Response.” *Review of General Psychology* 17 (1):1–19. URL `https://doi.org/10.1037/a0030997`.

Nail, Paul R. and Marilyn D. Van Leeuwen. 1993. "An Analysis and Restructuring of the Diamond Model of Social Response." *Pers Soc Psychol Bull* 19 (1):106–116. URL https://doi.org/10.1177/0146167293191012.

Nolan, Jessica M., P. Wesley Schultz, Robert B. Cialdini, Noah J. Goldstein, and Vladas Griskevicius. 2008. "Normative Social Influence is Underdetected." *Personality and Social Psychology Bulletin* 34 (7):913–923. URL https://doi.org/10.1177/0146167208316691. PMID: 18550863.

Nyczka, P., K. Byrka, P. R. Nail, and K. Sznajd-Weron. 2018. "Conformity in numbers - Does criticality in social responses exist?" *PLoS ONE* 13 (12).

Nyczka, Piotr and Katarzyna Sznajd-Weron. 2013. "Anticonformity or Independence? Insights from Statistical Physics." *Journal of Statistical Physics* 151 (1):174–202. URL https://doi.org/10.1007/s10955-013-0701-4.

Opper, S., V. Nee, and S. Brehm. 2015. "Homophily in the career mobility of China's political elite." *Social Science Research* 54:332–352.

Perez-Truglia, Ricardo and Guillermo Cruces. 2017. "Partisan Interactions: Evidence from a Field Experiment in the United States." *Journal of Political Economy* 125 (4):1208–1243. URL https://doi.org/10.1086/692711.

Rendell, L., R. Boyd, D. Cownden, M. Enquist, K. Eriksson, M. W. Feldman, L. Fogarty, S. Ghirlanda, T. Lillicrap, and K. N. Laland. 2010. "Why Copy Others? Insights from the Social Learning Strategies Tournament." *Science* 328 (5975):208–213. URL https://science.sciencemag.org/content/sci/328/5975/208.full.pdf.

Riach, P. A. and J. Rich. 2002. "Field Experiments of Discrimination in the Market Place." *The Economic Journal* 112 (483):F480–F518. URL https://doi.org/10.1111/1468-0297.00080.

Robin, Stéphane, Agnieszka Rusinowska, and Marie Claire Villeval. 2014. "Ingratiation: Experimental evidence." *European Economic Review* 66:16 – 38. URL http://www.sciencedirect.com/science/article/pii/S0014292113001414.

Sakha, Sahra and Antonia Grohmann. 2016. "The Effect of Peer Observation on Consumption Choices: Experimental Evidence." Tech. rep., Bundesbank Discussion Paper No. 01/2016. Available at SSRN: https://ssrn.com/abstract=2797074.

Schaerf, Timothy M., Peter W. Dillingham, and Ashley J. W. Ward. 2017. "The effects of external cues on individual and collective behavior of shoaling fish." *Science Advances* 3 (6). URL https://advances.sciencemag.org/content/3/6/e1603201.

Schlag, Karl H. 1998. "Why Imitate, and If So, How?: A Boundedly Rational Approach to Multi-armed Bandits." *Journal of Economic Theory* 78 (1):130–156.

Schlag, Karl H. 1999. "Which one should I imitate?" *Journal of Mathematical Economics* 31 (4):493–522.

Schmelz, Katrin. 2011. "E-nstructions: A tool for electronic instructions in laboratory experiments." Jena Economic Research Papers, No. 2011-008.

Schultz, P. Wesley, Jessica M. Nolan, Robert B. Cialdini, Noah J. Goldstein, and Vladas Griskevicius. 2007. "The Constructive, Destructive, and Reconstructive Power of Social Norms." *Psychological Science* 18 (5):429–434.

Schumpeter, Joseph Alois. 1934. *The Theory of Economic Development*. Cambridge, MA: Harvard University Press.

Schwarz, Gideon. 1978. "Estimating the Dimension of a Model." *Ann. Statist.* 6 (2):461–464. URL https://doi.org/10.1214/aos/1176344136.

Selten, R. 1967. *Beiträge zur experimentellen Wirtschaftsforschung*, chap. Die Strategiemethode zur Erforschung des eingeschränkt rationalen Verhaltens im Rahmen eines Oligopolexperiments. Mohr, Tübingen, 136–168.

Shalley, Christina E. and Jill E. Perry-Smith. 2001. "Effects of Social-Psychological Factors on Creative Performance: The Role of Informational and Controlling Expected Evaluation and Modeling Experience." *Organizational Behavior and Human Decision Processes* 84 (1):1 – 22. URL http://www.sciencedirect.com/science/article/pii/S0749597800929182.

Shane, S. A. 1992. "Why do some societies invent more than others." *Journal of Business Venturing* 7 (1):29–46.

Siedlecki, Patryk, Janusz Szwabinski, and Weron Tomasz. 2016. "The Interplay Between Conformity and Anticonformity and its Polarizing Effect on Society." *Journal of Artificial Societies and Social Simulation* 19 (4).

Simon, Herbert A. 1955. "A Behavioral Model of Rational Choice." *The Quarterly Journal of Economics* 69 (1):99–118.

Simpson, J. A., S. W. Gangestad, P. N. Christensen, and K. Leck. 1999. "Fluctuating asymmetry, sociosexuality, and intrasexual competitive tactics." *Journal of Personality and Social Psychology* 76 (1):159–172. URL `<GotoISI>://WOS:000078323600013`.

Smith, Lones and Peter Sorensen. 2000. "Pathological Outcomes of Observational Learning." *Econometrica* 68 (2):371–398. URL `https://onlinelibrary.wiley.com/doi/abs/10.1111/1468-0262.00113`.

Smith, Sarah, Frank Windmeijer, and Edmund Wright. 2015. "Peer Effects in Charitable Giving: Evidence from the (Running) Field." *The Economic Journal* 125 (585):1053–1071.

Sosna, Matthew M. G., Colin R. Twomey, Joseph Bak-Coleman, Winnie Poel, Bryan C. Daniels, Pawel Romanczuk, and Iain D. Couzin. 2019. "Individual and collective encoding of risk in animal groups." *Proceedings of the National Academy of Sciences* 116 (41):20556–20561. URL `https://www.pnas.org/content/116/41/20556`.

Tajfel, Henri. 1970. "Experiments in intergroup discrimination." *Scientific American* 223 (5):96–103.

Tibbetts, E. A. and J. Dale. 2007. "Individual recognition: it is good to be different." *Trends in ecology & evolution* 22 (10):529–537.

Touboul, Jonathan. 2019. "The hipster effect: When anti-conformists all look the same." *Discrete & Continuous Dynamical Systems - B* 24 (8):4379–4415. URL `http://dx.doi.org/10.3934/dcdsb.2019124`.

Turner, John C, Rupert J Brown, and Henri Tajfel. 1979. "Social comparison and group interest in ingroup favouritism." *European journal of social psychology* 9 (2):187–204.

Tversky, Amos. 1969. "Intransitivity of preferences." *Psychological Review* 76 (1):31–48.

Vega-Redondo, Fernando. 1997. "The Evolution of Walrasian Behavior." *Econometrica* 65 (2):375–384.

Vine, Ian. 1971. "Risk of visual detection and pursuit by a predator and the selective advantage of flocking behaviour." *Journal of Theoretical Biology* 30 (2):405 – 422. URL `http://www.sciencedirect.com/science/article/pii/0022519371900610`.

Viscido, Steven V. and David S. Wethey. 2002. “Quantitative analysis of fiddler crab flock movement: evidence for selfish herd behaviour.” *Animal Behaviour* 63 (4):735 – 741. URL http://www.sciencedirect.com/science/article/pii/S0003347201919359.

Vives, Xavier. 1997. “Learning from Others: A Welfare Analysis.” *Games and Economic Behavior* 20 (2):177 – 200. URL http://www.sciencedirect.com/science/article/pii/S0899825697905625.

Wiessner, P. 2002. “Hunting, healing, and hxaro exchange - A long-term perspective on !Kung (Ju/’hoansi) large-game hunting.” *Evolution and Human Behavior* 23 (6):407–436. URL <GotoISI>://WOS:000179057500001.

Willis, R. H. and J. M. Levine. 1976. *Social psychology: An introduction*, chap. Interpersonal influence and conformity. New York: Free Press, 309–341.

Willis, Richard H. 1963. “Two Dimensions of Conformity-Nonconformity.” *Sociometry* 26 (4):499–513. URL http://www.jstor.org/stable/2786152.

———. 1965. “Conformity, Independence, and Anticonformity.” *Human Relations* 18 (4):373–388. URL https://doi.org/10.1177/001872676501800406.

Wright, Daniel B., Kamala London, and Michael Waechter. 2009. “Social anxiety moderates memory conformity in adolescents.” *Applied Cognitive Psychology* 24 (7):1034–1045. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/acp.1604.

Yamagishi, Toshio, Hirofumi Hashimoto, and Joanna Schug. 2008. “Preferences Versus Strategies as Explanations for Culture-Specific Behavior.” *Psychological Science* 19 (6):579–584. URL https://doi.org/10.1111/j.1467-9280.2008.02126.x. PMID: 18578848.

A Theoretical Appendix

A.1 Responses to evaluation based on salience

Proposition 1 (Salience-based punishment)**.** *The A players follow their own taste. If the A players both disagree with B then B chooses as the A players if $\frac{\tau_B}{m} < \frac{2}{3}$, and B is indifferent if $\frac{\tau_B}{m} = \frac{2}{3}$. In all other cases, B follows his own taste.*

Proof. First, we study the behavior of player B. If the A players disagree then B will not be punished independent of the own choice because there is always an A player who is salient. Thus, B follows the own taste. If B agrees with both A players, then B has no incentive not to follow the own taste. Deviating from the own taste would increase the punishment probability from $\frac{1}{3}$ to 1. If B disagrees with both A players, then following the own taste provides a utility of $\tau_B - m$ and adjusting to the choice of the A players provides a utility of $-\frac{m}{3}$. Thus, B strictly prefers the own taste if $\tau_B - m > -\frac{m}{3} \Leftrightarrow \frac{\tau_B}{m} > \frac{2}{3}$.

Without loss of generality, we can use A_1 in order to study the behavior of the A players. We show that A_1 has a lower probability to be punished by following the own taste, independent of the strategies of A_2 and B. In Table 8, we show the punishment probability for A_1 for the strategies F =follow the own taste, S =switch to non-favorite taste, C =conformity (only possible for B), i.e. adjust to As if they agree. For convenience, we define $q = 1 - p$.

Table 8: Punishment probabilities based on salience

A_1				F	S	F	S	F	S	F	S
A_2				F	F	S	S	F	F	S	S
B				F	F	F	F	C	C	C	C
A_1	A_2	B	Probability								
X	X	X	p^3	$\frac{1}{3}$	1	0	0	$\frac{1}{3}$	1	0	$\frac{1}{3}$
X	X	Y	p^2q	0	0	1	$\frac{1}{3}$	$\frac{1}{3}$	0	1	$\frac{1}{3}$
X	Y	X	p^2q	0	0	$\frac{1}{3}$	1	0	$\frac{1}{3}$	$\frac{1}{3}$	1
Y	X	X	p^2q	1	$\frac{1}{3}$	0	0	1	$\frac{1}{3}$	$\frac{1}{3}$	0
X	Y	Y	pq^2	1	$\frac{1}{3}$	0	0	1	$\frac{1}{3}$	$\frac{1}{3}$	0
Y	X	Y	pq^2	0	0	$\frac{1}{3}$	1	0	$\frac{1}{3}$	$\frac{1}{3}$	1
Y	Y	X	pq^2	0	0	1	$\frac{1}{3}$	$\frac{1}{3}$	0	1	$\frac{1}{3}$
Y	Y	Y	q^3	$\frac{1}{3}$	1	0	0	$\frac{1}{3}$	1	0	$\frac{1}{3}$

Punishment probabilities for A_1 depending on the preferred options (X vs. Y) and strategies (F, S or C).

The difference of the probability to be punished when switching vs. when not provides the following expressions:

$$p^3(1-\frac{1}{3})+p^2q(\frac{1}{3}-1)+pq^2(\frac{1}{3}-1)+q^3(1-\frac{1}{3})=\frac{2}{3}(p^2-q^2)(p-q)>0$$

$$p^2q(\frac{4}{3}-\frac{4}{3})+pq^2(\frac{4}{3}-\frac{4}{3})=0$$

$$p^3(1-\frac{1}{3})+p^2q(-\frac{1}{3}+\frac{1}{3}+\frac{1}{3}-1)+pq^2(\frac{1}{3}-1+\frac{1}{3}-\frac{1}{3})+q^3(1-\frac{1}{3})=\frac{2}{3}(p^2-q^2)(p-q)>0$$

$$p^3(\frac{1}{3})+p^2q(\frac{1}{3}-1+1-\frac{1}{3}-\frac{1}{3})+pq^2(-\frac{1}{3}+1-\frac{1}{3}+\frac{1}{3}-1)+q^3(\frac{1}{3})=\frac{1}{3}(p^2-q^2)(p-q)>0$$

All expressions are weakly positive. Since we assumed that $\tau_{A_1} > 0$, A_1 has a strict incentive to follow the own taste. □

Proposition 2 (Salience-based reward)**.** *There is a unique symmetric equilibrium, which is characterized as follows. The A players choose according to their own taste if* $\frac{\tau_A}{m} > K$ *where K is a constant that depends on T, and p. They choose against their preferred taste if* $\frac{\tau_A}{m} < K$*, and are indifferent if* $\frac{\tau_A}{m} = K$*. If the A players both agree with B then B chooses contrary to the A players if* $\frac{\tau_B}{m} < \frac{2}{3}$*, and is indifferent if* $\frac{\tau_B}{m} = \frac{2}{3}$*. In all other cases, B chooses according to his own taste.*

Proof salience-based reward. First, we study the behavior of player B. If the A players disagree then B will not be rewarded independent of his own choice. Thus, B follows the own taste. If B disagrees with both A players, then B follows his own taste. Deviating from the own taste would decrease the reward probability from 1 to $\frac{1}{3}$. If B agrees with both A players, then following his own taste provides a utility of $\tau_B + \frac{m}{3}$ and switching to the choice not made by the A players provides a utility of m. Thus, B strictly prefers the own taste if $\tau_B + \frac{m}{3} > m \Leftrightarrow \frac{\tau_B}{m} > \frac{2}{3}$.

Let us now turn to the incentives for player A_1. Let φ be the share of the player B who follows the own taste, i.e. the players with $\tau_B > \frac{2}{3}m$. Since T is strictly increasing φ is well defined. We now determine the probability γ_i that A_i chooses according to the own taste.

We show the reward probabilities of A_1 in Table 8. Again, we define $q = 1 - p$.

The difference of the reward probability between when A_1 follows the own taste and when it does not equals

$$p^3\gamma_2(\frac{\varphi}{3}-1)+p^2q\gamma_2(1-\frac{\varphi}{3})+pq^2\gamma_2(1-\frac{\varphi}{3})+q^3\gamma_2(\frac{\varphi}{3}-1)=$$

$$-(1-\frac{\varphi}{3})\gamma_2(p^3-p^2q-pq^2+q^3)=$$

$$-(1-\frac{\varphi}{3})\gamma_2(p-q)^2$$

Table 9: Reward probabilities based on salience

A_1				F	S	F	S
A_2				F	F	S	S
A_1	A_2	B	Probability				
X	X	X	p^3	$\frac{\varphi}{3}$	1	0	0
X	X	Y	p^2q	0	0	1	$\frac{\varphi}{3}$
X	Y	X	p^2q	0	0	$\frac{\varphi}{3}$	1
Y	X	X	p^2q	1	$\frac{\varphi}{3}$	0	0
X	Y	Y	pq^2	1	$\frac{\varphi}{3}$	0	0
Y	X	Y	pq^2	0	0	$\frac{\varphi}{3}$	1
Y	Y	X	pq^2	0	0	1	$\frac{\varphi}{3}$
Y	Y	Y	q^3	$\frac{\varphi}{3}$	1	0	0

Reward probabilities for A_1 depending on the preferred options (X vs. Y) and strategies (F, S or C).

Thus, A_1 follows the own taste if $\tau > \gamma_2 m(1 - \frac{\varphi}{3})(p-q)^2$. Let τ_{crit} be the threshold above which the A player follow their own taste. Then the share of players who follow the own taste equals $1 - T(\tau_{crit})$ and we get the following equation.

$$\gamma_1 = 1 - T(\gamma_2 m(1 - \frac{\varphi}{3})(2p-1)^2)$$

If we set $\gamma_1 = \gamma_2 =: \gamma$, we get a unique solution for γ because we assumed T to be continuous. The share γ decreases in m and in p. If T shifts to the left (people care less about the own taste), then γ decreases. This is the case because the direct effect and the indirect effect via φ go into the same direction. □

There can be asymmetic equilibria, also if the distribution is uniform. Assume, that T is uniformly distributed between 1 and $\frac{1}{\sigma}$, i.e. $T(\gamma) = \sigma\gamma$. $K := m(1 - \frac{\varphi}{3}(2p-1)^2$. Then

$$\begin{aligned} \gamma_1 &= 1 - \sigma K \gamma_2 \\ \gamma_2 &= 1 - \sigma K \gamma_1 \\ \gamma_1 &= 1 - \sigma K(1 - \sigma K \gamma_1) \\ \gamma_1(1 - (\sigma K)^2) &= 1 - \sigma K \end{aligned}$$

If $\sigma K = 1$ then any combination with $\gamma_1 + \gamma_2 = 1$ is an equilibrium. Otherwise, there is only the symmetric equilibrium with $\gamma_1 = \gamma_2 = \frac{1}{1+\sigma K}$.

A.2 Responses to evaluation based on homophily

Proposition 3 (Homophily-based punishment)**.** *Independent of the strategy of player B, the A players always follow their own taste. B is conformist if $\frac{\tau_B}{m} < (2(p-\frac{1}{2})^2+\frac{1}{6})$, otherwise B is independent. (In case of equality B is indifferent between conformity and independence.)*

Proposition 4 (Homophily-based reward)**.** *Independent of the strategy of player B, the A players always follow their own taste. B is conformist if $\frac{\tau_B}{m} < (\frac{1}{3} - 2p(1-p))$. B is anticonformist if $\frac{\tau_B}{m} < (\frac{2}{3} - \frac{p^4+(1-p)^4}{p^3+(1-p)^3})$. (In case of equality B is indifferent between conformity or anticonformity and independence.)*

Proof. We first show that A_1 has an incentive to follow the own taste independent of the strategies of A_2 and B, and in both the reward and the punishment setting. To do this, we setup Table tab:homophily-punishment. It contains the difference between the probability that A_1 gets punished when switching compared to following the own taste. A positive entry corresponds to an incentive to follow the own taste. For all combinations of strategies of B and A_2, this difference is a homogeneous polynomial in p and $q = 1-p$ of grade 4. The coefficients of these polynomials can be found in the rows $M40$ to $M04$.

It can be shown that all these polynomials are positive. Note that all polynomials are symmetric in the sense that the coefficient of $p^k q^{1-k}$ equals the coefficient of $p^{1-k}q^k$. Most of the terms have the form $p^k q^{1-k} - p^r q^{1-r} - p^{1-r}q^r + p^{1-k}q^k$ with $k > r \geq 2$. In this case, we get

$$\begin{aligned} p^k q^{1-k} - p^r q^{1-r} - p^{1-r}q^r + p^{1-k}q^k &= \\ p^r q^{1-k}(p^{k-r} - q^{k-r}) - p^{1-k}q^r(p^{k-r} - q^{k-r}) &= \\ (p^r q^{1-k} - p^{1-k}q^r)(p^{k-r} - q^{k-r}) &> 0 \end{aligned}$$

This argument works for column 1, 2, 8, 11, 15, and 16. In columns 3, 4, 5, 7, 9 10,12, and 13 the polynomials can be composed into the sum of two polynomials of this form. For example, polynomial 3 can be decomposed as follows:

$$\begin{aligned} &p^4q^0 - \frac{1}{2}p^3q^2 - p^2q^2 - \frac{1}{2}p^1q^3 + p^0q^4 = \\ &\qquad \frac{1}{2}(p^4q^0 - p^3q^1 - p^1q^3 + p^0q^4) + \frac{1}{2}(p^4q^0 - p^2q^2 - p^2q^2 + p^0q^4) \end{aligned}$$

Table 10: Punishment probability differences based on homophily

				(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)
			*A*_2	F	S	F	S	F	S	F	S	F	S	F	S	F	S	F	S
			B0	F	F	F	F	F	F	F	F	S	S	S	S	S	S	S	S
			B1	F	F	F	F	S	S	S	S	F	F	F	F	S	S	S	S
			B2	F	F	S	S	F	F	S	S	F	F	S	S	F	F	S	S
Eval	*A*_1	*A*_2	*B*																
X	X	X	X	$\frac{2}{3}$	$\frac{1}{2}$	1	$\frac{1}{2}$	$\frac{1}{6}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{2}{3}$	$\frac{1}{3}$	1	$\frac{1}{3}$	$\frac{1}{6}$	$\frac{1}{3}$	$\frac{1}{2}$	$\frac{1}{3}$
X	X	X	Y	$\frac{1}{2}$	$\frac{1}{3}$	$\frac{1}{2}$	$\frac{1}{2}$	1	$\frac{1}{3}$	1	$\frac{1}{2}$	$\frac{1}{6}$	$\frac{1}{3}$	$\frac{1}{6}$	$\frac{1}{2}$	$\frac{2}{3}$	$\frac{1}{3}$	$\frac{2}{3}$	$\frac{1}{2}$
X	X	Y	X	$\frac{1}{2}$	$\frac{2}{3}$	$\frac{1}{2}$	1	$\frac{1}{2}$	$\frac{1}{6}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{3}$	$\frac{2}{3}$	$\frac{1}{3}$	1	$\frac{1}{3}$	$\frac{1}{6}$	$\frac{1}{3}$	$\frac{1}{2}$
X	Y	X	X	$-\frac{2}{3}$	$-\frac{1}{2}$	-1	$-\frac{1}{2}$	$-\frac{1}{6}$	$-\frac{1}{2}$	$-\frac{1}{2}$	$-\frac{1}{2}$	$-\frac{2}{3}$	$-\frac{1}{3}$	-1	$-\frac{1}{3}$	$-\frac{1}{6}$	$-\frac{1}{3}$	$-\frac{1}{2}$	$-\frac{1}{3}$
Y	X	X	X	$-\frac{1}{3}$	$-\frac{1}{2}$	$-\frac{1}{2}$	$-\frac{1}{2}$	$-\frac{1}{3}$	-1	$-\frac{1}{2}$	-1	$-\frac{1}{3}$	$-\frac{1}{6}$	$-\frac{1}{2}$	$-\frac{1}{6}$	$-\frac{1}{3}$	$-\frac{2}{3}$	$-\frac{1}{2}$	$-\frac{2}{3}$
X	X	Y	Y	$\frac{1}{3}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{3}$	1	$\frac{1}{2}$	1	$\frac{1}{3}$	$\frac{1}{6}$	$\frac{1}{2}$	$\frac{1}{6}$	$\frac{1}{3}$	$\frac{2}{3}$	$\frac{1}{2}$	$\frac{2}{3}$
X	Y	X	Y	$-\frac{1}{2}$	$-\frac{1}{3}$	$-\frac{1}{2}$	$-\frac{1}{2}$	-1	$-\frac{1}{3}$	-1	$-\frac{1}{2}$	$-\frac{1}{6}$	$-\frac{1}{3}$	$-\frac{1}{6}$	$-\frac{1}{2}$	$-\frac{2}{3}$	$-\frac{1}{3}$	$-\frac{2}{3}$	$-\frac{1}{2}$
X	Y	Y	X	$-\frac{1}{2}$	$-\frac{2}{3}$	$-\frac{1}{2}$	-1	$-\frac{1}{2}$	$-\frac{1}{6}$	$-\frac{1}{2}$	$-\frac{1}{2}$	$-\frac{1}{3}$	$-\frac{2}{3}$	$-\frac{1}{3}$	-1	$-\frac{1}{3}$	$-\frac{1}{6}$	$-\frac{1}{3}$	$-\frac{1}{2}$
Y	X	X	Y	$-\frac{1}{2}$	$-\frac{2}{3}$	$-\frac{1}{2}$	-1	$-\frac{1}{2}$	$-\frac{1}{6}$	$-\frac{1}{2}$	$-\frac{1}{2}$	$-\frac{1}{3}$	$-\frac{2}{3}$	$-\frac{1}{3}$	-1	$-\frac{1}{3}$	$-\frac{1}{6}$	$-\frac{1}{3}$	$-\frac{1}{2}$
Y	X	Y	X	$-\frac{1}{2}$	$-\frac{1}{3}$	$-\frac{1}{2}$	$-\frac{1}{2}$	-1	$-\frac{1}{3}$	-1	$-\frac{1}{2}$	$-\frac{1}{6}$	$-\frac{1}{3}$	$-\frac{1}{6}$	$-\frac{1}{2}$	$-\frac{2}{3}$	$-\frac{1}{3}$	$-\frac{2}{3}$	$-\frac{1}{2}$
Y	Y	X	X	$\frac{1}{3}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{3}$	1	$\frac{1}{2}$	1	$\frac{1}{3}$	$\frac{1}{6}$	$\frac{1}{2}$	$\frac{1}{6}$	$\frac{1}{3}$	$\frac{2}{3}$	$\frac{1}{2}$	$\frac{2}{3}$
X	Y	Y	Y	$-\frac{1}{3}$	$-\frac{1}{2}$	$-\frac{1}{2}$	$-\frac{1}{2}$	$-\frac{1}{3}$	-1	$-\frac{1}{2}$	-1	$-\frac{1}{3}$	$-\frac{1}{6}$	$-\frac{1}{2}$	$-\frac{1}{6}$	$-\frac{1}{3}$	$-\frac{2}{3}$	$-\frac{1}{2}$	$-\frac{2}{3}$
Y	X	Y	Y	$-\frac{2}{3}$	$-\frac{1}{2}$	-1	$-\frac{1}{2}$	$-\frac{1}{6}$	$-\frac{1}{2}$	$-\frac{1}{2}$	$-\frac{1}{2}$	$-\frac{2}{3}$	$-\frac{1}{3}$	-1	$-\frac{1}{3}$	$-\frac{1}{6}$	$-\frac{1}{3}$	$-\frac{1}{2}$	$-\frac{1}{3}$
Y	Y	X	Y	$\frac{1}{2}$	$\frac{2}{3}$	$\frac{1}{2}$	1	$\frac{1}{2}$	$\frac{1}{6}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{3}$	$\frac{2}{3}$	$\frac{1}{3}$	1	$\frac{1}{3}$	$\frac{1}{6}$	$\frac{1}{3}$	$\frac{1}{2}$
Y	Y	Y	X	$\frac{1}{2}$	$\frac{1}{3}$	$\frac{1}{2}$	$\frac{1}{2}$	1	$\frac{1}{3}$	1	$\frac{1}{2}$	$\frac{1}{6}$	$\frac{1}{3}$	$\frac{1}{6}$	$\frac{1}{2}$	$\frac{2}{3}$	$\frac{1}{3}$	$\frac{2}{3}$	$\frac{1}{2}$
Y	Y	Y	Y	$\frac{2}{3}$	$\frac{1}{2}$	1	$\frac{1}{2}$	$\frac{1}{6}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{2}{3}$	$\frac{1}{3}$	1	$\frac{1}{3}$	$\frac{1}{6}$	$\frac{1}{3}$	$\frac{1}{2}$	$\frac{1}{3}$
			M40	$\frac{2}{3}$	$\frac{1}{2}$	1	$\frac{1}{2}$	$\frac{1}{6}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{2}{3}$	$\frac{1}{3}$	1	$\frac{1}{3}$	$\frac{1}{6}$	$\frac{1}{3}$	$\frac{1}{2}$	$\frac{1}{3}$
			M31	0	0	$-\frac{1}{2}$	$\frac{1}{2}$	1	-1	$\frac{1}{2}$	$-\frac{1}{2}$	$-\frac{1}{2}$	$\frac{1}{2}$	-1	1	$\frac{1}{2}$	$-\frac{1}{2}$	0	0
			M22	$-1\frac{1}{3}$	-1	-1	-2	$-2\frac{1}{3}$	1	-2	0	$-\frac{1}{3}$	$-1\frac{2}{3}$	0	$-2\frac{2}{3}$	$-1\frac{1}{3}$	$\frac{1}{3}$	-1	$-\frac{2}{3}$
			M13	0	0	$-\frac{1}{2}$	$\frac{1}{2}$	1	-1	$\frac{1}{2}$	$-\frac{1}{2}$	$-\frac{1}{2}$	$\frac{1}{2}$	-1	1	$\frac{1}{2}$	$-\frac{1}{2}$	0	0
			M04	$\frac{2}{3}$	$\frac{1}{2}$	1	$\frac{1}{2}$	$\frac{1}{6}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{2}{3}$	$\frac{1}{3}$	1	$\frac{1}{3}$	$\frac{1}{6}$	$\frac{1}{3}$	$\frac{1}{2}$	$\frac{1}{3}$

Probability differences between switching and following for player A_1 for all strategies of player A_2 and B. Positive values means that switching increases the punishment probability. The columns starting with Mxy contain the coefficient of the monomial $p^x(1-p)^y$.

For columns 6 and 14, we need a calculation as the following illustrating the case 6:

$$\begin{aligned}\frac{1}{2}p^4 - p^3q + p^2q^2 - pq^3 + \frac{1}{2}q^4 &= \\ \frac{1}{4}(p-q)^4 + \frac{1}{4}p^4 - (\frac{6}{4}-1)p^2q^2 + \frac{1}{4}q^4 &= \\ -\frac{1}{4}(p-q)^4 + \frac{1}{4}(p^2-q^2)^2 &< 0\end{aligned}$$

Table 11 shows the corresponding table for reward. In this table we show the difference in the reward probability between following and switching. Thus also in this case, positive coefficients support following the own taste. The arguments why the polynomials are positive are analogous to the arguments in the punishment case. Thus, the A players have a monetary incentive to follow their taste, in addition to their direct incentive τ_i.

Now, we determine the strategies for player B.

Punishment case. If B does not disagree with both A players it is best to follow the own taste. So, the only relevant case is that B disagrees with both As. In this case, the probability that B prefers the same option as the evaluator equals $\frac{2p^2q^2}{p^3q+2p^2q^2+pq^3} = 2pq$. Thus, following the own taste provides a utility of $\tau_B - (1-2p(1-p))m$. Switching the choice provides a utility of $-\frac{m}{3}$. Thus, B is conform if $\tau_B < m(1-2p(1-p)-\frac{1}{3})$.

Reward case. If the A players disagree, there is no reason for B to deviate from the own taste. The probability that the evaluator has the same taste as B is at least $\frac{1}{2}$ and if B chooses according to the taste of the evaluator, the winning probability is $\frac{1}{2}$ independent of the choice. If B disagrees with both As then the probability that B prefers the same option as the evaluator equals $2p(1-p)$ (as above). Thus, B is conform if $\frac{m}{3} > \tau_B + 2p(1-p)m \Leftrightarrow \tau_B < m(\frac{1}{3} - 2p(1-p))$. If B agrees with the As then the probability that B prefers the same option as the evaluator equals $\frac{p^4+(1-p)^4}{p^4+p^3(1-p)+p(1-p)^3+(1-p)^4} = \frac{p^4+(1-p)^4}{p^3+(1-p)^3}$. Thus, B is anticonform if $m(1-\frac{p^4+(1-p)^4}{p^3+(1-p)^3}) > \tau_B + m\frac{1}{3} \Leftrightarrow \tau_B < m(\frac{2}{3} - \frac{p^4+(1-p)^4}{p^3+(1-p)^3})$. □

A.3 Responses to evaluations based on performance

The generally preferred option may also be interpreted as the correct option, in particular in the domain of objective facts. In this case, evaluators could potentially reward and punish based on performance, i.e., they could punish someone who took a decision that is probably wrong and reward a choice that is probably true. This relates to the information cascade literature (Banerjee, 1992; Bikhchandani, Hirshleifer, and Welch, 1992; Anderson and Holt, 1997), but the relation is not very tight because only player B can be part of a cascade.[25]

[25] Our setting specifically relates to Guarino, Harmgart, and Huck (2011). They do not provide information on the choice sequence, which is comparable to our situation where the evaluator remains uninformed of who

Table 11: Reward probability differences based on homophily

			A_2	F	S	F	S	F	S	F	S	F	S	F	S	F	S	F	S
			B0	F	F	F	F	F	F	F	F	S	S	S	S	S	S	S	S
			B1	F	F	F	F	S	S	S	S	F	F	F	F	S	S	S	S
			B2	F	F	S	S	F	F	S	S	F	F	S	S	F	F	S	S
Eval	A_1	A_2	*B*																
X	X	X	X	$\frac{1}{3}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{3}$	1	$\frac{1}{2}$	1	$\frac{1}{3}$	$\frac{1}{6}$	$\frac{1}{2}$	$\frac{1}{6}$	$\frac{1}{3}$	$\frac{2}{3}$	$\frac{1}{2}$	$\frac{2}{3}$
X	X	X	Y	$\frac{1}{2}$	$\frac{2}{3}$	$\frac{1}{2}$	1	$\frac{1}{2}$	$\frac{1}{6}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{3}$	$\frac{2}{3}$	$\frac{1}{3}$	1	$\frac{1}{3}$	$\frac{1}{6}$	$\frac{1}{3}$	$\frac{1}{2}$
X	X	Y	X	$\frac{1}{2}$	$\frac{1}{3}$	$\frac{1}{2}$	$\frac{1}{2}$	1	$\frac{1}{3}$	1	$\frac{1}{2}$	$\frac{1}{6}$	$\frac{1}{3}$	$\frac{1}{6}$	$\frac{1}{2}$	$\frac{2}{3}$	$\frac{1}{3}$	$\frac{2}{3}$	$\frac{1}{2}$
X	Y	X	X	$-\frac{1}{3}$	$-\frac{1}{2}$	$-\frac{1}{2}$	$-\frac{1}{2}$	$-\frac{1}{3}$	-1	$-\frac{1}{2}$	-1	$-\frac{1}{3}$	$-\frac{1}{6}$	$-\frac{1}{2}$	$-\frac{1}{6}$	$-\frac{1}{3}$	$-\frac{2}{3}$	$-\frac{1}{2}$	$-\frac{2}{3}$
Y	X	X	X	$-\frac{2}{3}$	$-\frac{1}{2}$	-1	$-\frac{1}{2}$	$-\frac{1}{6}$	$-\frac{1}{2}$	$-\frac{1}{2}$	$-\frac{1}{2}$	$-\frac{2}{3}$	$-\frac{1}{3}$	-1	$-\frac{1}{3}$	$-\frac{1}{6}$	$-\frac{1}{3}$	$-\frac{1}{2}$	$-\frac{1}{3}$
X	X	Y	Y	$\frac{2}{3}$	$\frac{1}{2}$	1	$\frac{1}{2}$	$\frac{1}{6}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{2}{3}$	$\frac{1}{3}$	1	$\frac{1}{3}$	$\frac{1}{6}$	$\frac{1}{3}$	$\frac{1}{2}$	$\frac{1}{3}$
X	Y	X	Y	$-\frac{1}{2}$	$-\frac{2}{3}$	$-\frac{1}{2}$	-1	$-\frac{1}{2}$	$-\frac{1}{6}$	$-\frac{1}{2}$	$-\frac{1}{2}$	$-\frac{1}{3}$	$-\frac{2}{3}$	$-\frac{1}{3}$	-1	$-\frac{1}{3}$	$-\frac{1}{6}$	$-\frac{1}{3}$	$-\frac{1}{2}$
X	Y	Y	X	$-\frac{1}{2}$	$-\frac{1}{3}$	$-\frac{1}{2}$	$-\frac{1}{2}$	-1	$-\frac{1}{3}$	-1	$-\frac{1}{2}$	$-\frac{1}{6}$	$-\frac{1}{3}$	$-\frac{1}{6}$	$-\frac{1}{2}$	$-\frac{2}{3}$	$-\frac{1}{3}$	$-\frac{2}{3}$	$-\frac{1}{2}$
Y	X	X	Y	$-\frac{1}{2}$	$-\frac{1}{3}$	$-\frac{1}{2}$	$-\frac{1}{2}$	-1	$-\frac{1}{3}$	-1	$-\frac{1}{2}$	$-\frac{1}{6}$	$-\frac{1}{3}$	$-\frac{1}{6}$	$-\frac{1}{2}$	$-\frac{2}{3}$	$-\frac{1}{3}$	$-\frac{2}{3}$	$-\frac{1}{2}$
Y	X	Y	X	$-\frac{1}{2}$	$-\frac{2}{3}$	$-\frac{1}{2}$	-1	$-\frac{1}{2}$	$-\frac{1}{6}$	$-\frac{1}{2}$	$-\frac{1}{2}$	$-\frac{1}{3}$	$-\frac{2}{3}$	$-\frac{1}{3}$	-1	$-\frac{1}{3}$	$-\frac{1}{6}$	$-\frac{1}{3}$	$-\frac{1}{2}$
Y	Y	X	X	$\frac{2}{3}$	$\frac{1}{2}$	1	$\frac{1}{2}$	$\frac{1}{6}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{2}{3}$	$\frac{1}{3}$	1	$\frac{1}{3}$	$\frac{1}{6}$	$\frac{1}{3}$	$\frac{1}{2}$	$\frac{1}{3}$
X	Y	Y	Y	$-\frac{2}{3}$	$-\frac{1}{2}$	-1	$-\frac{1}{2}$	$-\frac{1}{6}$	$-\frac{1}{2}$	$-\frac{1}{2}$	$-\frac{1}{2}$	$-\frac{2}{3}$	$-\frac{1}{3}$	-1	$-\frac{1}{3}$	$-\frac{1}{6}$	$-\frac{1}{3}$	$-\frac{1}{2}$	$-\frac{1}{3}$
Y	X	Y	Y	$-\frac{1}{3}$	$-\frac{1}{2}$	$-\frac{1}{2}$	$-\frac{1}{2}$	$-\frac{1}{3}$	-1	$-\frac{1}{2}$	-1	$-\frac{1}{3}$	$-\frac{1}{6}$	$-\frac{1}{2}$	$-\frac{1}{6}$	$-\frac{1}{3}$	$-\frac{2}{3}$	$-\frac{1}{2}$	$-\frac{2}{3}$
Y	Y	X	Y	$\frac{1}{2}$	$\frac{1}{3}$	$\frac{1}{2}$	$\frac{1}{2}$	1	$\frac{1}{3}$	1	$\frac{1}{2}$	$\frac{1}{6}$	$\frac{1}{3}$	$\frac{1}{6}$	$\frac{1}{2}$	$\frac{2}{3}$	$\frac{1}{3}$	$\frac{2}{3}$	$\frac{1}{2}$
Y	Y	Y	X	$\frac{1}{2}$	$\frac{2}{3}$	$\frac{1}{2}$	1	$\frac{1}{2}$	$\frac{1}{6}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{3}$	$\frac{2}{3}$	$\frac{1}{3}$	1	$\frac{1}{3}$	$\frac{1}{6}$	$\frac{1}{3}$	$\frac{1}{2}$
Y	Y	Y	Y	$\frac{1}{3}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{3}$	1	$\frac{1}{2}$	1	$\frac{1}{3}$	$\frac{1}{6}$	$\frac{1}{2}$	$\frac{1}{6}$	$\frac{1}{3}$	$\frac{2}{3}$	$\frac{1}{2}$	$\frac{2}{3}$
			M40	$\frac{1}{3}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{3}$	1	$\frac{1}{2}$	1	$\frac{1}{3}$	$\frac{1}{6}$	$\frac{1}{2}$	$\frac{1}{6}$	$\frac{1}{3}$	$\frac{2}{3}$	$\frac{1}{2}$	$\frac{2}{3}$
			M31	0	0	$-\frac{1}{2}$	$\frac{1}{2}$	1	-1	$\frac{1}{2}$	$-\frac{1}{2}$	$-\frac{1}{2}$	$\frac{1}{2}$	-1	1	$\frac{1}{2}$	$-\frac{1}{2}$	0	0
			M22	$-\frac{2}{3}$	-1	0	-2	$-2\frac{2}{3}$	0	-2	-1	$\frac{1}{3}$	$-1\frac{1}{3}$	1	$-2\frac{1}{3}$	$-1\frac{2}{3}$	$-\frac{1}{3}$	-1	$-1\frac{1}{3}$
			M13	0	0	$-\frac{1}{2}$	$\frac{1}{2}$	1	-1	$\frac{1}{2}$	$-\frac{1}{2}$	$-\frac{1}{2}$	$\frac{1}{2}$	-1	1	$\frac{1}{2}$	$-\frac{1}{2}$	0	0
			M04	$\frac{1}{3}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{3}$	1	$\frac{1}{2}$	1	$\frac{1}{3}$	$\frac{1}{6}$	$\frac{1}{2}$	$\frac{1}{6}$	$\frac{1}{3}$	$\frac{2}{3}$	$\frac{1}{2}$	$\frac{2}{3}$

Probability differences between following and switching for player A_1 for all strategies of player A_2 and B. Positive values means that following increases the reward probability. The columns starting with Mxy contain the coefficient of the monomial $p^x(1-p)^y$.

Interestingly, there are also equilibria in which "wrong signals" are sent. For example, if the evaluator favors the minority (punishes one of the majority or rewards the minority player), then the A players may have an incentive to choose the option they do not prefer. Because in this case the majority is not evidence for the better option, this can be an equilibrium. We present the equilibria in which the A players choose according the own taste. They always exist. We get the following propositions for the punishment treatment.

Proposition 5 (Performance-based punishment)**.** *The equilibria in which the A players choose according to their taste can be described with the parameter $\eta \in [0,1]$. The evaluator has a real choice only when the three group members disagree and a majority (two players) chooses one option and the minority (one player) chooses the other. If the group members disagree in their choice and the evaluator's taste matches the majority choice, then the evaluator punishes the minority player. If the group members disagree in their choice and the evaluator's taste contradicts the majority choice then the evaluator punishes one of the majority players with probability η and otherwise the minority player. B is conformist if $\frac{\tau_B}{m} > \frac{1}{3}(1-p^2-q^2) - 2p^2q^2(1-\eta)$, otherwise B is independent. (In case of equality B is indifferent between conformity and independence.)*

Proof. If the A players choose according their own taste, then the majority choice is at least as informative as the own taste of the evaluator. Accordingly, if the evaluator has the same taste as the majority, he will punish the minority. Thus, for B, anticonformity does not make sense because B will be punished with a probability of at least $\frac{1}{2}$ Let κ be the share of B who conform. Then the probability that the majority is correct equals

$$\begin{aligned}\frac{(1-\kappa)p^2q + 2p^2q}{(1-\kappa)p^2q + 2p^2q + 2p^2q + (1-\kappa)pq^2} &= \\ \frac{(kappa)p^2q}{(3-\kappa)p^2q + (3-\kappa)pq^2} &= \\ \frac{p}{p+q} &= p\end{aligned}$$

Thus, independent of the conformity of player B, the evaluator chooses with the majority when he agrees with it and is indifferent otherwise.

If B always follows the own taste the punishment probability of B equals $\frac{1}{3}$ because in this case all three players have the same strategy. If B is conform then the punishment probability of B equals $\frac{1}{3}(p^3+p^2q)+p^2q^2(1-\eta)+p^2q^2(1-\eta)+\frac{1}{3}(pq^2+q^3) = \frac{1}{3}(p^2+q^2)+2p^2q^2(1-\eta) \leq \frac{1}{3}$.

the B player is who could condition the choice. However, their setting differs in that the information is not symmetric in the options.

The difference of the punishing probability of B between when B follows the own taste and when B is conform equals $\frac{1}{3} - \frac{1}{3}(p^2+q^2)+2p^2q^2(1-\eta) = \frac{1}{3}(1-p^2-q^2)-2p^2q^2(1-\eta) > 0$. Thus, B follows the own taste if $\frac{\tau_B}{m} > \frac{1}{3}(1-p^2-q^2) - 2p^2q^2(1-\eta)$.

It remains to be shown that following the own taste is an equilibrium for the A players. When facing different decisions of the A players, player B cannot affect the probability of reward because he will be in the majority anyhow. Thus, player B chooses according to the own taste because $\tau > 0$. We show that A_1 has an incentive to follow the own taste if the evaluator chooses according to the own taste or according to the majority, and if B does not choose against his taste when the A players disagree in their choice. If the evaluator follows the own taste, this has been shown in Table 10 above. If the evaluator goes with the majority (and punishes the minority, we get the polynomials $\frac{2}{3}p^4 - 1\frac{1}{3}p^2q^2 + \frac{2}{3}q^4$, $p^4 - 2p^2q^2 + q^4$, $\frac{2}{3}p^4 - 1\frac{1}{3}^2q^2 + \frac{2}{3}q^4$, and $p^4 - 2p^2q^2 + q^4$ for when B follows the own taste, is disconform, conform and does not choose according to the own taste when the A players agree in their choice. These polynomials are positive. □

For the reward treatment, we get the following propositions.

Proposition 6 (Performance-based reward)**.** *The equilibria in which the A players choose according to their taste can be described by the parameter* $\eta \in [0, \frac{1}{3pq} - (p^2+q^2)]$. *The evaluator has a real choice only when the three group members disagree and a majority (two players) chooses one option and the minority (one player) chooses the other. If the group members disagree in their choice and the evaluator's taste matches the majority choice, then the evaluator rewards one of the majority players. If the group members disagree in their choice and the evaluator's taste contradicts the majority choice, then the evaluator rewards one of the minority player with probability* η *and otherwise one of the majority player. The A players always follow their taste. B is conformist if* $\frac{\tau_B}{m} > \frac{1}{3}(p^2+q^2) + (p^3q+pq^3)(1-\eta) - \frac{1}{3}$, *otherwise B is independent. (In case of equality B is indifferent between conformity or anticonformity and independence.)*

Proof. Let κ be the share of Bs who conform and α be the share of Bs who anticonform. Then, the probability that the majority is correct equals

$$
\begin{aligned}
\frac{\alpha p^3 + (1-\kappa)p^2q + 2p^2q}{\alpha p^3 + (1-\kappa)p^2q + 2p^2q + 2p^2q + (1-\kappa)pq^2 + \alpha q^3} &= \\
\frac{\alpha p^3 + (3-\kappa)p^2q}{\alpha p^3 + (kappa)p^2q + (3-\kappa)pq^2 + \alpha q^3} &\geq \\
\frac{\alpha p^3 + (3-\kappa)p^2q}{\alpha p^3 + (3-\kappa)p^2q + (3-\kappa)p^2q + \alpha pq^2} &= \\
\frac{p^2}{p^2 + pq} &= p
\end{aligned}
$$

Equality holds if $\alpha = 0$. If there is anticonformity, then the evaluator has a strict incentive to reward someone of the majority. However, in this case anticonformity prevents from getting the reward and will not be applied. Thus, there is no equilibrium with anticonformity.

If B always follows the own taste, the reward probability of B equals $\frac{1}{3}$ because in this case all three players have the same strategy. If B is anticonform then the reward probability of B equals $p^3q\eta + p^2q^2\eta + p^3q + pq^3 + p^2q^2\eta + pq^3\eta$. This expression is larger than $\frac{1}{3}$ if $\eta > \frac{\frac{1}{3} - p^3q - pq^3}{p^3q + 2p^2q^2 + pq^3} = \frac{1}{3pq} - (p^2 + q^2)$. Thus, an equilibrium exists only if $\eta \leq \frac{1}{3pq} - (p^2 + q^2)$.

If B conforms then the reward probability of B equals $\frac{1}{3}(p^3 + p^2q) + p^3q(1-\eta) + pq^3(1-\eta) + \frac{1}{3}(pq^2 + q^3) = \frac{1}{3}(p^2 + q^2) + (p^3q + pq^3)(1-\eta)$. Thus, B is conform if $\frac{\tau_B}{m} > \frac{1}{3}(p^2 + q^2) + (p^3q + pq^3)(1-\eta) - \frac{1}{3}$.

The proof that following the own taste is an equilibrium for the A players is done as in the prove above. □

A.4 Larger groups

In an earlier version of this paper, we analyzed equilibria for the case $N > 3$ but with monetary incentive only. The results are comparable. There is conformity in the punishment treatment for salience-based punishment as well as for homophily-based punishment. In these situations, the A players follow their taste. There is (trivially) anticonformity in salience based reward and the A players randomize in this case. Most complicated is the case of homophily-based reward. We can show that conformity is more likely for a higher p and anticonformity is more likely for a lower p. The A players follow their own taste. The last result could only be shown for $N \leq 1000$, though.

B Statistical Appendix

B.1 Similarity of colors

Figure 8 illustrates three distance variables we use to quantify the similarity of a target color to the three colors of the other group members (Colors 1-3). The different values of the distance variables are graphically illustrated by the total length of the red segments in each panel. Figure 8 illustrates the three distance variables for the RGB color metric. The RGB metric measures the difference between two colors by their Euclidean distance in the intensity of the three basic components red, green and blue.

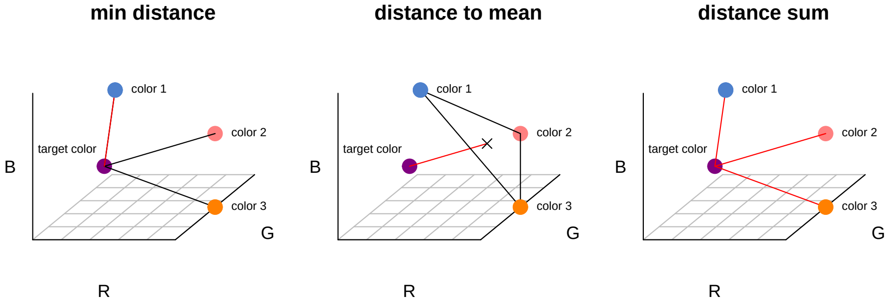


Figure 8: Distance variables for the similarity of colors

Red lines in the three panels illustrate the three variables used to determine color similarity in Experiment 2 for the RGB metric. The axes labels R, G, and B indicate the three components red, green, and blue of the RGB colorspace. The black and red lines between colors reflect the Euclidean distance between two colors in the three-dimensional space. *min distance* is the minimum of the three Euclidean distances, *distance to mean* the Euclidean distance to the average color, and *distance sum* the sum of the three Euclidean distances.

The variable *min distance* in the left panel reflects the minimum of the three Euclidean distances of the chosen color to each of the colors chosen by the group members in the RGB color space. The variable *distance to mean* in the central panel reflects the Euclidean distance to the average of the other three colors. The variable *distance sum* in the right panel reflects the sum of the three Euclidean distances.[26]

[26] While the RGB color metric is the most common specification to measure color distance, it does not effectively reflect perceived differences in colors. Therefore, we additionally calculate the values of the three variables illustrated in Figure 8 for the ΔE^* distance metric. The distance metric ΔE^* was proposed by the International Commission on Illumination in 1976 to eliminate perceptual non-linearity in the RGB colorspace. It has been refined twice to better fit the human perception of differences in color. We use the R package *colorscience* (Gama and Davis, 2018) to generate color differences based on the most recent definition of the ΔE^* metric (CIE 2000).

For the statistical analyses, we mainly focus on the *min distance* variable, which has been proposed as a measure of the spatial cohesion of individuals in groups (Clark and Evans, 1954). We use the *min distance* variable in combination with the RGB metric since the RGB metric is the most frequently used color difference metric. We check the robustness of the experimental results of Experiment 2 based on the two remaining distance variables.

B.2 Operationalization

To calculate coordinates of the response to social influence for the binary choice data of Experiment 1, we focus on all informed choices in which the participant is informed that both other group members prefer the same alternative. We assume that the unanimity of choices of the other group members exerts social influence. We assume that social influence is not exerted if the choices of the other group members diverge. To calculate the coordinates of the social response in the model space, we use formulae (7) and (8) together with a simple distance function that is positive if the informed choice differs from the alternative chosen by both other group members and that is zero otherwise.

For Experiment 2, we elicit participants' uninformed choices which are subsequently transmitted to the other members of the group. Each group member makes her informed choice after being informed about the uninformed choices of the other group members. Participants know that the informed choice of one randomly selected participant would be evaluated together with the uninformed choices of the other group members. To calculate coordinates of the response to social information for the multinomial choice data of Experiment 2, we focus on all situations in which the informed choice could be adjusted in both directions, towards and away from the behavior of the other group members. We use formulae (3) and (4) to calculate the coordinates of the social response in the model space.

B.2.1 Multinomial choices

To fix ideas, assume we observe N pairs of nonsocial and informed choices with index $i = \{1, \ldots, N\}$. Let Δ be a variable that indicates the difference of a choice to the choices of others. Let Δ_i^{ns} and Δ_i^{s} indicate the values of this variable for the uninformed choice and the informed choice of the ith pair of choices. We define the probability to observe an adjustment of the informed choice *towards* the choices of others as the relative frequency of adjustments that decrease Δ:

$$P(towards) = \frac{\Sigma_{i=1}^{N} I(\Delta_i^{ns} > \Delta_i^{s})}{N} \tag{1}$$

We define the probability to observe an adjustment *away* from the choices of others as the relative frequency of adjustments that increase Δ:

$$P(away) = \frac{\Sigma_{i=1}^{N} I(\Delta_i^{ns} < \Delta_i^{s})}{N} \tag{2}$$

The coordinates (x, y) which locate the observed response to social influence in the model space are:

$$x = P(towards) + P(away) \tag{3}$$

$$y = P(towards) - P(away) \tag{4}$$

Two comments are in order. First, the operationalization neglects the size of the adjustment $|\Delta_i^{ns} - \Delta_i^{s}|$ which may contain information about the response to social influence. Second, formulas (1) and (2) assume that it is possible to adjust every informed choice in both directions. This is usually the case in our experimental setup.

B.2.2 Binary choices

If the choice format is binary it will only be possible to adjust in one of the two directions. In this case, we estimate $P_b(towards)$ by the relative frequency of adjustment for observations N^t in which an adjustment of the informed choice *towards* the choices of others is possible. We estimate $P_b(away)$ by the relative frequency of adjustment for observations N^a in which an adjustment of the informed choice *away from* the choices of others is possible. The corresponding equations are:

$$P_b(towards) = \frac{\Sigma_{i \in N^t} I(\Delta_i^{ns} > \Delta_i^{s})}{|N^t|} \tag{5}$$

and

$$P_b(away) = \frac{\Sigma_{i \in N^a} I(\Delta_i^{ns} < \Delta_i^{s})}{|N^a|} \tag{6}$$

The coordinates (x, y) which locate the binary choices in the model space are:

$$x = P_b(towards) + P_b(away) \tag{7}$$

$$y = P_b(towards) - P_b(away) \tag{8}$$

The coordinates yield an estimate for the location of the response to social influence under the assumption that adjustments in each direction are possible in half of the observations. This might not be true given the data but yields an unbiased estimate of the location of the response to social influence.

B.3 Supplementary figures and tables

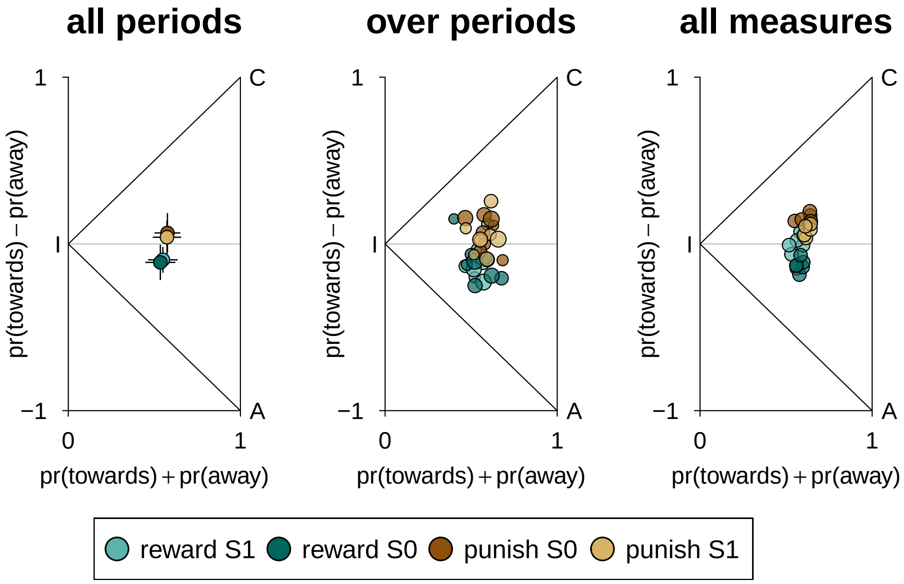


Figure 9: Robustness of treatment effect in Experiment 2

Left panel: average response to social influence for data of all periods. Central panel: evolution of the average response effect over periods. Bigger dots reflect later periods. Right panel: average response for each of the 6 possible combinations of the three distance variables and the two color metrics.

Table 12: Average distance of adjusted choices across treatments

	reward	punishment	t-statistic	df	p-value
S0 & S1 POOLED					
RGB distance					
min distance	0.56	0.50	3.60	85	<0.001
sum distances	2.37	2.23	2.67	85	0.005
distance to mean	0.66	0.61	2.57	85	0.006
rank min distance	2.68	2.41	4.51	82	<0.001
rank sum distances	2.65	2.39	4.00	82	<0.001
rank distance to mean	2.62	2.40	3.41	82	<0.001
ΔE^* distance (CIE, 2000)					
min distance	30.87	26.73	4.50	85	<0.001
sum distances	139.2	129.5	2.93	83	0.002
distance to mean	40.19	36.90	2.67	78	0.005
rank min distance	2.72	2.43	4.21	83	<0.001
rank sum distances	2.66	2.39	4.08	82	<0.001
rank distance to mean	2.60	2.41	2.73	81	0.004
S0 TREATMENTS					
RGB distance					
min distance	0.56	0.5	2.79	36	0.004
sum distances	2.38	2.23	2.07	38	0.023
distance to mean	0.67	0.61	2.32	39	0.013
rank min distance	2.74	2.42	3.88	43	<0.001
rank sum distances	2.72	2.35	4.7	45	<0.001
rank distance to mean	2.7	2.34	4.44	45	<0.001
ΔE^* distance (CIE, 2000)					
min distance	32.06	26.75	4.19	42	<0.001
sum distances	143.85	130.13	3.16	37	0.002
distance to mean	42.34	37.38	3.01	34	0.002
rank min distance	2.78	2.43	3.65	44	<0.001
rank sum distances	2.76	2.39	4.27	40	<0.001
rank distance to mean	2.71	2.43	3.07	41	0.002
S1 TREATMENTS					
RGB distance					
min distance	0.55	0.5	2.25	34	0.015
sum distance	2.37	2.22	1.67	35	0.052
distance to mean	0.65	0.61	1.29	36	0.102
rank min distance	2.61	2.4	2.45	37	0.010
rank sum distances	2.56	2.44	1.2	33	0.120
rank distance to mean	2.52	2.46	0.6	30	0.276
ΔE^* distance (CIE, 2000)					
min distance	29.5	26.71	2.13	32	0.020
sum distances	133.9	128.82	1.03	35	0.155
distance to mean	37.71	36.32	0.8	36	0.214
rank min distance	2.64	2.43	2.23	38	0.016
rank sum distances	2.53	2.4	1.47	37	0.074
rank distance to mean	2.47	2.39	0.76	38	0.225

Table 12 shows averages of matching group averages.

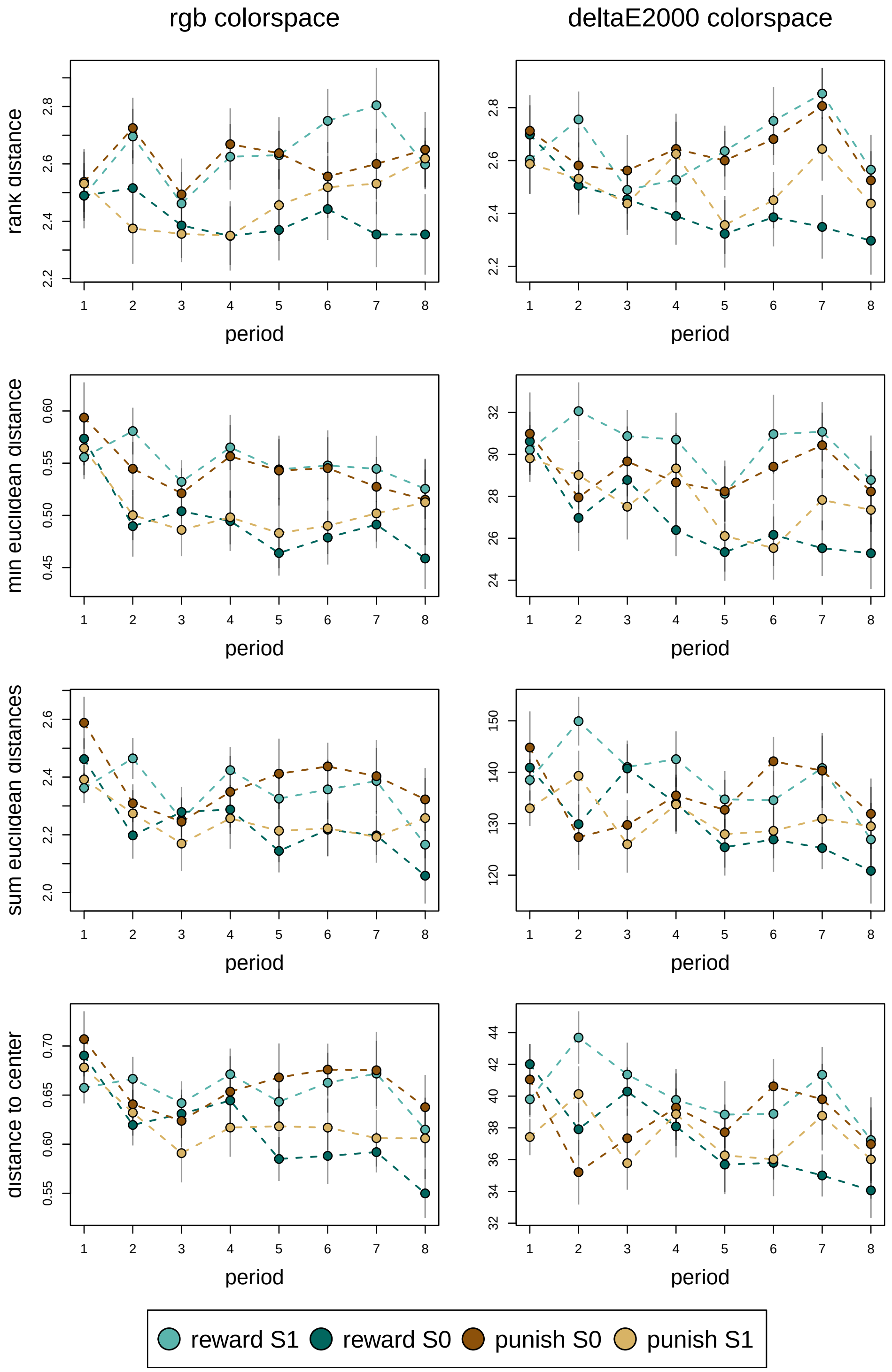


Figure 10: Evolution of distance measures over periods

Evolution of the average distance of informed choices over the eight periods of Experiment 2. The dots indicate period specific means of the six distance variables and the rank of the minimal distance in combination with the rgb metric. Whiskers indicate plus/minus one standard error of the mean, based on 10000 block bootstrap samples (group ID).

B.4 Analysis of heterogeneity

To analyze heterogeneity in conditional choices, we fit mixture models with K response types to the data of each treatment and select K based on the Bayesian information criterion (Schwarz, 1978). We use the R package *stratEst* (Dvorak, 2023) to obtain maximum likelihood estimates and block-bootstrapped standard errors of the parameters of the mixture models. The log likelihood of the mixture model is

$$\ln L = \sum_{i=1}^{N} \ln \left(\sum_{k=1}^{K} p_k \prod_{s=1}^{S} \prod_{r=1}^{R} (\pi_{ksr})^{y_{isr}} \right). \tag{9}$$

where p_k denotes the frequency of type k in the sample, s is an index for the choice situations the participants $i \in \{1, \cdots, N\}$ are confronted with in the experiment, r the number of alternatives in these situations, and y_{isr} the number of times participant i shows response r in situation s.

For the data of the first experiment $S = 2$ applies as we focus on two situations, one in which conformity is possible and the other in which anticonformity is possible. In both situations $R = 2$ applies as there are only two responses possible: adjust or not. For the data of the second experiment $S = 1$ and $R = 3$ applies as we focus exclusively on the situation where an adjustment in the direction of conformity, an adjustment in the direction of anticonformity, and no adjustment are possible.

Estimates and standard errors of type position

Let π_k^t and π_k^a be the maximum likelihood estimates of the probabilities that type k adjusts *towards* and *away* from others' choices respectively.

For Experiment 1, $\pi_k^t = \pi_{ks\prime r\prime}$ where $s\prime$ indicates the situation in which conformity is possible and $r\prime$ the response to adjust. $\pi_k^a = \pi_{ks^* r\prime}$ where s^* indicates the situation in which anticonformity is possible.

For Experiment 2, $\pi_k^t = \pi_{ksr\prime}$ where $r\prime$ indicates the response to adjust in the direction of conformity, and $\pi_k^a = \pi_{ksr^*}$ where r^{star} indicates the response to adjust n the direction of anticonformity.

The coordinates of type k in the two dimensional model space are calculated based on:

$$x_k = \pi_k^t + \pi_k^a \quad \text{and} \quad y_k = \pi_k^t - \pi_k^a.$$

The standard errors of the coordinates se_{x_k} and se_{y_k} are estimated by block-bootstrapping the variance-covariance matrix of the response probabilities π_{ksr}:

$$se_{x_k} = \sqrt{var(\pi_k^t) + var(\pi_k^a) + 2cov(\pi_k^t, \pi_k^a)}$$

$$se_{y_k} = \sqrt{var(\pi_k^t) + var(\pi_k^a) - 2cov(\pi_k^t, \pi_k^a)}$$

where $var(\cdot)$ and $cov(\cdot, \cdot)$ denote the entries corresponding to the response probabilities in the block-bootstrapped variance-covariance matrix.

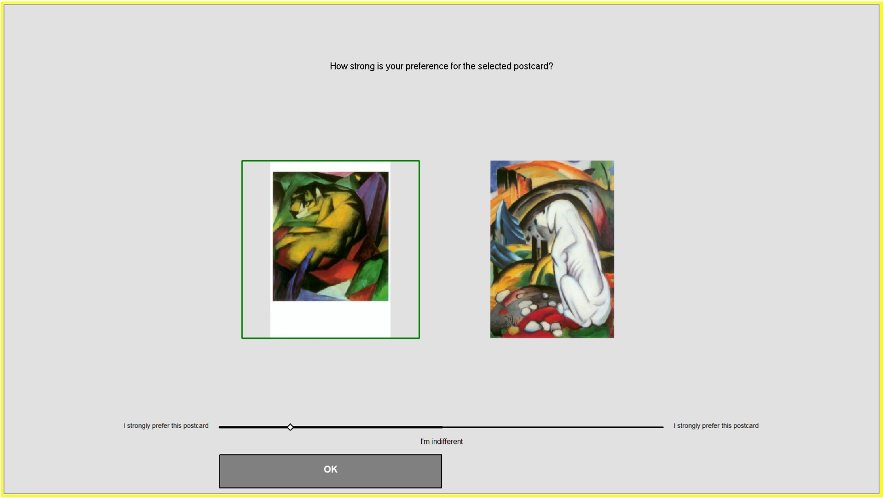
C SUPPLEMENTARY MATERIAL

C.1 The Python Code

The following Python code provides a check for the incentive for A to follow the own taste. If this is not the case, a message is displayed.

```
import math
import scipy.special
from scipy.stats import binom
from fractions import Fraction
import time

# this is the f restriction
def oddsF(N,k):
    H = int((N-1)/2)
    if (k == 0):
        return Fraction(N - 1, 1)
    if (k > 0 and k < H):
        return Fraction(N - k, k + 1)
    if (k == H):
        return Fraction(1, 1)
    if (k > H and k < N - 1):
        return Fraction(k + 1, N -k)
    if (k == N - 1):
        return Fraction(N - 1, 1)
    return Fraction(1, 1)


def Coeff_k_1_p2(N, k, betak):
    if betak == 0:
        return +Fraction( scipy.special.comb(N-2, k-1, exact=True), k+1)
    else:
        return +Fraction( scipy.special.comb(N-2, k-1, exact=True), k)

def Coeff_k_p1(N, k, betak):
    if betak == 0:
        return -Fraction( scipy.special.comb(N-2, k, exact=True), N-k-1)
    else:
        return -Fraction( scipy.special.comb(N-2, k, exact=True), N-k)

def Coeff_Nk2_p1(N, k, betak):
    if betak == 0:
        return +Fraction( scipy.special.comb(N-2, k, exact=True), N-k-1)
    else:
        return +Fraction( scipy.special.comb(N-2, k, exact=True), N-k)

def Coeff_Nk1_p0(N, k, betak):
    if betak == 0:
        return -Fraction( scipy.special.comb(N-2, k-1, exact=True), k+1)
    else:
        return -Fraction( scipy.special.comb(N-2, k-1, exact=True), k)


def GetStepR( N, R, f, k, b):
    OK = True
    if R >= 0:
         R = Fraction(0,1)
    if k == N-1:
        if f == Fraction(N-1, 1):
            R = R*Fraction(1, N) + Fraction(1, N)
        else:
            R = R/f + Fraction(1, N-1)
    else:
        A = Coeff_k_1_p2(N, k, b)
        B = Coeff_k_p1(N, k, b)
        if A*f <= -R:
            print("ERROR␣CheckHigh(N,k0),␣A*f<␣-R:", N, R, f, k, b,A,B)
            OK = False
        R = A + B + R/f
    return OK, R

def GetStepS( N, S, f, k, b):
    OK = True
    if S < 0:
        OK = False
        print("ERROR:␣GetStepS(S,␣f,␣k,␣b):,␣S␣<␣0:", N, S, f, k, b)
    if k == N-1 :
        if f == Fraction(N-1, 1):
            S = S*Fraction(N-1, N) - Fraction(1, N)
        else:
            S = S*f - Fraction(1, N-1)
    else:
        C=Coeff_Nk2_p1(N, k, b)
        D=Coeff_Nk1_p0(N, k, b)
        if f*S >= -D:
            S = C + D/f + S
        else:
            S = C + D + f*S
    return OK, S


def CheckLow(N):
    H = int((N-1)/2)
    NIsOdd = H*2 == N-1

    for k0 in range(0, H):
        f = oddsF(N, k0)
        R = Fraction(0, 1)
        S = Fraction(0, 1)
        Dp2 = Fraction(0, 1)
        Dp0 = Fraction(0, 1)
        if (k0 == 0):
            R = -Fraction(1, N)
            S = Fraction(1, N)
            Dp2 = -Fraction(1, N)
            Dp0 = Fraction(1, N)
        else:
            R = -Fraction(1, N-1)
            S = Fraction(1, N-1)
            Dp2 = Fraction(0, 1)
            Dp0 = Fraction(0, 1)

        for k in range(1, H):

            if R >= 0:
                 R = Fraction(0, 1)

            A = Coeff_k_1_p2(N, k, 0)
            B = Coeff_k_p1(N, k, 0)
            if (k0 == 0 and k == 1): A = A + Dp2
            if A*f <= -R:
                print("CheckLow(N,k0),␣A*f␣<␣-R", N, k0, beta, k, R, A, B)
                return False
            R0 = A + B + R/f

            A = Coeff_k_1_p2(N, k, 1)
            B = Coeff_k_p1(N, k, 1)
            if (k0 == 0 and k == 1): A = A + Dp2
            if A*f <= -R:
                print("CheckLow(N,k0),␣A*f␣<␣-R:", N, k0, beta, k, R, A, B
                   )
                return False
            R1 = A + B + R/f
            if k < k0:       # for k < k0 we use beta = 0
                R = R0
            elif k == k0:    # for k = k0 we use beta = 1
                R = R1
            else:            # for k > k0 we use the more conservative
               estimation
                if R0 < R1: R = R0
                else: R = R1

            C = Coeff_Nk2_p1(N, k, 0)
            D = Coeff_Nk1_p0(N, k, 0)
            if (k0 == 0 and k == 1): D = D + Dp0
            if S < 0:
                print("CheckLow(N,k0),␣S␣<␣0:", N, k0, k, f, S, C, D)
                return False
            if f*S >= -D:
                S0 = C + D/f + S
            else:
                S0 = C + D + f*S

            C = Coeff_Nk2_p1(N, k, 1)
            D = Coeff_Nk1_p0(N, k, 1)
            if (k0 == 0 and k == 1): D = D + Dp0
            if S < 0:
                print("CheckLow(N,k0),␣S␣<␣0:", N, k0, k, f, S, C, D)
                return False
            if f*S >= -D:
                S1 = C + D/f + S
            else:
                S1 = C + D + f*S

            if k < k0:       # for k < k0 we use beta = 0
                S = S0
            elif k == k0:    # for k = k0 we use beta = 1
                S = S1
            else:            # for k > k0 we use the more conservative
               estimation
                if S0 < S1: S = S0
                else: S = S1

        if (not NIsOdd):    # we check both at H-1 and at H
            ok = S >= 0 and S+R >= 0 # OK because the ther term of H is
               positive
            if  not ok:
                stepRok, R = GetStepR(N, R, f, H, 0)
                stepSok, S  = GetStepS(N, S, f, H, 0)
                if (not(stepRok and stepSok)):
                    return False
                ok = S >= 0 and S + R >= 0
            if not ok:
                print("CheckLow(N,k0)␣even:", ok, N, k0, R, S, R+S)
                return False
        else:
            A = Coeff_k_1_p2(N, H, 0)
            B = Coeff_k_p1(N, H, 0)
            C = Coeff_Nk2_p1(N, H, 0)
            D = Coeff_Nk1_p0(N, H, 0)
            RH  = R + f*A + C
            SH = S + B + D/f
            ok = RH >= 0 and RH >= -SH
            if not ok:
                print("CheckLow(N,k0)␣odd:", ok, N, k0, RH, SH, RH + SH)
                return False
    return True

def CheckHigh(N):
    H = int((N-1)/2)
    NIsOdd = H*2 == N-1
    for L in range(H+1, N):
        f = oddsF(N, L)
        R = Fraction(0, 1)
        S = Fraction(0, 1)

        for k in range(H+1, L+1):
            stepRok, R = GetStepR(N, R, f, k, 0)
            stepSok, S  = GetStepS(N, S, f, k, 0)
            if (not(stepRok and stepSok)):
                return False

        k= L
        while  k < N-1 and R < 0:
            k += 1
            stepRok, R = GetStepR(N, R, f, k, 1)
            stepSok, S  = GetStepS(N, S, f, k, 1)
            if (not(stepRok and stepSok)):
                return False

        if not( R >= 0 and S >=0 ):
            print("ERROR␣CheckHigh(N):", N, L, R, S)
            return False
    return True

def Check(N):
    ok = CheckHigh(N)
    if ok: ok = CheckLow(N)
    return ok

N = 4
state = "all␣ok"
EndTime = time.time()
InitialTime = EndTime
while (N <= 1000):
    StartTime = EndTime
    ok = Check(N)
    if (not ok): state = "not␣all␣ok"
    EndTime = time.time()
    print(ok, N, round(EndTime-StartTime, 1), "s,␣", round( (EndTime-
       InitialTime)/3600, 2), "h,␣", state)
    N += 1
```

C.2 Study materials

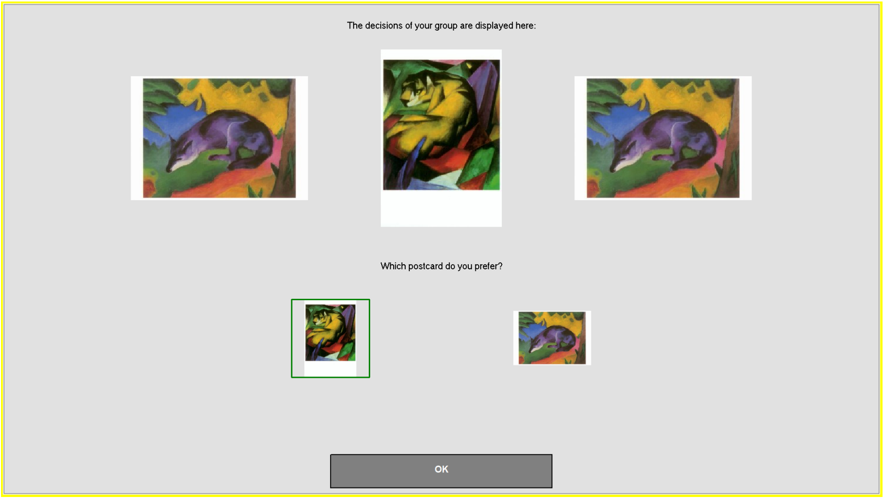


How strong is your preference for the selected postcard?

I strongly prefer this postcard
I strongly prefer this postcard
I'm indifferent
OK

Figure 11: Decision Screen of a related choice

Related choice comparing the one alternative of the informed choice (the tiger) to the common third alternative (dog). After participants select one option and confirm their selection, the slider in the lower part of the screen appears.

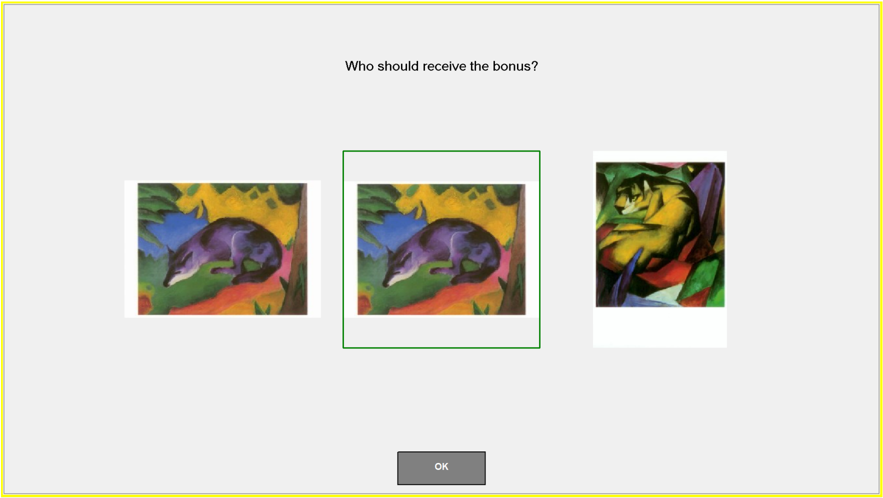


The decisions of your group are displayed here:

Which postcard do you prefer?

Figure 12: Decision screen of a informed choice

informed choice (tiger vs. fox). The decisions of the two other group members are depicted as the paintings on the left and right in the top line. The painting in the middle represents the choice currently selected by the participant.



Who should receive the bonus?

Figure 13: Decision screen of evaluator

Screen of an evaluation decision. The evaluator selects one of the three group members by clicking on one of the paintings. The evaluation decision has to be confirmed by clicking on the ”Ok” button.

Table 13: List of paintings

Set	Theme	Painting 1	Painting 2
1	Marc Chagall	Bride and Groom	The couple
2	Lyonel Feininger	The Grain Tower at Treptow on the Rega	Village Pond of Gelmeroda
3	Claude Monet	The Artist's Garden in Giverny	The artist's Garden in Giverney
4	Wassily Kandinsky	Improvisation 26	Improvisation 28 (2nd version)
5	August Macke	Garden Restaurant	Large Bright Walk
6	Franz Marc	Yellow Cow	The Dog in Front of the World
7	Caspar D. Friedrich	The Churchyard Gate	Hutten's Grave (Ruin of a Church Choir)
8	Houses	Egon Schiele, House with drying Laundry	Albrecht Duerer, The Castle of Trient
9	Work	Paul Cézanne, The Mowers	Vincent van Gogh, Field with Farmer and Mill
10	Women	August Macke, Portrait with Apples	Pablo Picasso, The Absinthe-Drinker
11	Trees	Paul Cézanne, The Chestnut Trees at Jas de Bouffan	Rudolph von Alt, Landscape in the Prater in Vienna
12	Hands	Albrecht Duerer, The Hands of Jesus Christ	Pablo Picasso, Crossed Hands
13	Ships	Egon Schiele, Fishing Boats in Trieste	Berthe Morisot, The Harbour of Nice
14	Flowers	Paul Cézanne, Flowers in a Vase and Fruit	Vincent van Gogh, Bouquet of Irises
15	Bridges	Claude Monet, Bridge over the Seine near Argenteuil	Vincent van Gogh, The Bridges of Asnières
Set	Theme	Painting 3	Painting 4
1	Marc Chagall	The Newly-Married of the Eiffel-Tower	Lovers
2	Lyonel Feininger	Gelmeroda IX	The Church of Halle
3	Claude Monet	The artist's Garden in Vétheuil	Resting under the Lilac
4	Wassily Kandinsky	Improvisation 34 (Orient II)	Improvisation Gorge
5	August Macke	Girls under Trees	Sunny Path
6	Franz Marc	The Tiger	Fox (Blue Black Fox)
7	Caspar D. Friedrich	Ruins of the Monastery Eldena	The Graveyard Door (The Churchyard)
8	Houses	Rudolph von Alt, The "Goldene Dachl" in Innsbruck	Carl Spitzweg, A Hypochondriac
9	Work	Carl Spitzweg, The Walk of the Boarding School	Pond at the Forest
10	Women	Egon Schiele, Peasants-Girl	Gustav Klimt, Johanna Staude
11	Trees	Vincent van Gogh, Road with Cypress and Star	Alfred Sisley, The Path to the Old Ferry at By
12	Hands	Egon Schiele, Clasped Hands	August Rodin, The Cathedral - Hands
13	Ships	Redon Odilon, The Mystical Boat	Raoul Dufy, The old Harbour of Marseille
14	Flowers	Claude Monet, Vase of Flowers	Pierre-Auguste Renoir, Bouquet of Chrysanthemums
15	Bridges	Alfred Sisley, Bridge near Hampton Court	William Turner, Old Welsh Bridge, Shrewsbury

Before each session, we selected 10 sets and for each set three of the paintings listed based on availability. Postcards of the paintings were ordered from Kunstverlag Reisser, Braunschweigstrasse 12, 1130 Vienna, Austria. Names are translations from German and taken from http://www.reisser-kunstpostkarten.de.

Table 14: Question sets 1-9

Set	Question	Option 1	Option 2	Option 3	Answer 1	Answer 2	Answer 3
1	Which country is larger (2015, m$\hat{2}$)?	Canada	USA	China	9984k	9826k	9596k
1	Which country is larger (2015, m$\hat{2}$)?	Portugal	Czech Republic	Austria	92090	78867	83871
1	Which country is larger (2015, m$\hat{2}$)?	Estonia	Denmark	Netherlands	45228	43094	41543
1	Which country is larger (2015, m$\hat{2}$)?	Lithuania	Croatia	Latvia	65300	56594	64589
1	Which country is larger (2015, m$\hat{2}$)?	Sudan	Indonesia	Mexico	1,861k	1,904k	1,964k
2	Which country has more inhabitants (2014)?	France	Italy	UK	65,835k	60,782k	64,351k
2	Which country has more inhabitants (2014)?	Spain	Ukraine	Poland	46,512k	45,245k	38,017k
2	Which country has more inhabitants (2014)?	Greece	Belgium	Czech Republic	10,926k	11,203k	10,512k
2	Which country has more inhabitants (2014)?	Austria	Switzerland	Bulgaria	8,506k	8,139k	7,245k
2	Which country has more inhabitants (2014)?	Malta	Luxemburg	Iceland	425k	549k	325k
3	Which company had more employees (2014)?	Bosch	Daimler	Metro	290,183	279,972	24,9150
3	Which company had more employees (2014)?	Bayer	ThyssenKrupp	Continental	118,900	160,745	189,168
3	Which company had more employees (2014)?	Lufthansa	BASF	BMW	118,781	113,292	116,324
3	Which company had more employees (2014)?	RWE	E.ON	MAN	59,784	58,503	55,903
3	Which company had more employees (2014)?	Bertelsmann	SAP	TUI	112,037	74,406	77,309
4	Who was born earlier?	Konrad Adenauer	F.D. Roosevelt	Theodor Heuss	1876	1882	1884
4	Who was born earlier?	Willy Brandt	John F. Kennedy	Walter Scheel	1913	1917	1919
4	Who was born earlier?	Helmut Schmidt	Richard Nixon	R. Weizsaecker	1918	1913	1920
4	Who was born earlier?	Horst Koehler	Gerhard Schroeder	Bill Clinton	1943	1944	1946
5	Which harbor is bigger (2014, TEU)?	Shanghai	Hong Kong	Singapore	35.3	22.30	33.9
5	Which harbor is bigger (2014, TEU)?	Hamburg	Antwerp	Los Angeles	9.7	9	8.3
5	Which harbor is bigger (2014, TEU)?	Guangzhou	Dubai	Rotterdam	16.2	15.2	12.3
6	Which airline had more passengers?	United Airlines	American Airlines	Ryanair	90,440k	87,830k	86,370k
6	Which airline had more passengers?	Lufthansa	Easyjet	Air China	59,850k	62,310k	54,580k
6	Which airline had more passengers?	Air Berlin	Brithish Airlines	Air France	29,910k	41,160k	45,410k
6	Which airline had more passengers?	KLM	Aeroflot	SAS	27,740k	23,600k	27,390k
7	Which country discharges more CO2 (2010, pp)?	Germany	Netherlands	Austria	12.3	10.1	12.1
7	Which country discharges more CO2 (2010, pp)?	Poland	Slovakia	Hungary	7.7	7.8	7.3
7	Which country discharges more CO2 (2010, pp)?	Lithuania	Latvia	Estonia	5.9	6.5	13.5
7	Which country discharges more CO2 (2010, pp)?	France	Portugal	Spain	9	6.9	8.5
7	Which country discharges more CO2 (2010, pp)?	Finland	Norway	Sweden	18.7	10.1	9.3
8	Which country has more inequality (2012, GINI)?	France	Belgium	Austria	33.1	27.6	30.5
8	Which country has more inequality (2012, GINI)?	Norway	Finland	Sweden	25.9	27.1	27.3
8	Which country has more inequality (2012, GINI)?	Bolivia	Ecuador	Peru	46.7	46.6	45.1
8	Which country has more inequality (2012, GINI)?	Costa Rica	Brazil	Argentina	48.6	52.7	42.5
8	Which country has more inequality (2012, GINI)?	Thailand	Laos	Vietnam	39.3	37.9	38.7
9	Which soccer club is worth more (2016)?	Manchester City	FC Chelsea	M. United	501.75	490	411.25
9	Which soccer club is worth more (2016)?	AS Rom	FC Valencia	SSC Neapel	250.7	282	284
9	Which soccer club is worth more (2016)?	Bayer 04 Leverkusen	VfL Wolfsburg	FC Schalke 04	211.1	183.1	199.8
9	Which soccer club is worth more (2016)?	Zenit St.Petersburg	AC Mailand	FC Sevilla	198.6	188.1	186.2

Sources: http://appsso.eurostat.ec.europa.eu, https://de.wikipedia.org, http://de.statista.com, http://carbonfootprintofnations.com, http://databank.worldbank.org, http://www.transfermarkt.de

Table 15: Question sets 10-19

Set	Question	Option 1	Option 2	Option 3	Answer 1	Answer 2	Answer 3
10	Which country won more medals (2014 Olympics?	Netherlands	France	Germany	24	15	19
10	Which country won more medals (2014 Olympics?	Switzerland	Sweden	Austria	11	15	17
10	Which country won more medals (2014 Olympics?	Canada	Norway	USA	26	28	
10	Which country won more medals (2014 Olympics?	Finland	UK	Ukraine	5	4	2
10	Which country won more medals (2014 Olympics?	Belarus	Kazakhstan	Australia	6	1	3
11	Which airport has more passengers (2014)?	Atlanta Int	L Heathrow	Dubai Int	96,178k	73,408k	70,475k
11	Which airport has more passengers (2014)?	Singapore Changi	Kuala Lumpur	Shanghai Int	54,093,000	48,930k	51,687k
11	Which airport has more passengers (2014)?	Charles de Gaulles	Frankfurt	A Schiphol	63,813k	59,566k	54,978k
11	Which airport has more passengers (2014)?	Madrid Barajas	SP-Guarulhos	Miami Int	41,822k	39,765k	40,941k
12	Who sold more records in Germany?	The Beatles	Michael Jackson	Madonna	7,600k	11,275k	12,300k
12	Who sold more records in Germany?	ACDC	ABBA	R. Williams	10,475k	10,800k	9,275k
12	Who sold more records in Germany?	Helene Fischer	Pur	Die Aerzte	9,150k	9,425k	7,850k
12	Who sold more records in Germany?	Britney Spears	Bon Jovi	Xavier Naidoo	5,050k	5,150k	5,525k
13	In which language is the letter ”a” more frequent?	German	English	French	6.51	8.167	7.636
13	In which language is the letter ”a” more frequent?	Spanish	Italian	Swedish	12.53	11.740	9.300
13	In which language is the letter ”a” more frequent?	German	English	French	17.4	12.702	14.715
13	In which language is the letter ”a” more frequent?	Spanish	Italian	Swedish	13.68	11.790	9.900
13	In which language is the letter ”a” more frequent?	Spanish	Italian	Swedish	6.71	6.880	8.800
14	Which initial letter is more common in German?	E	I	W	7.8	7.1	6.8
14	Which initial letter is more common in German?	H	I	O	7.232	6.286	6.264
14	Which initial letter is more common in German?	C	D	F	3.511	2.670	3.779
14	Which initial letter is more common in German?	J	K	V	0.597	0.590	0.649
15	Which country has more prisoners (2016, per 100k)?	USA	Cuba	Seychelles	698	510	799
15	Which country has more prisoners (2016, per 100k)?	Thailand	Russia	Ruanda	468	447	434
15	Which country has more prisoners (2015, absolute)?	Berlin	Saxony	Rhineland	3806	3385	3102
15	Which country has more prisoners (2015, absolute)?	Saxony-Anhalt	Thuringia	Hamburg	1670	1600	1559
15	Which country has more prisoners (2015, absolute)?	Schleswig-Holstein	Mecklenburg	Brandenburg	1162	1057	1324
16	Which food has more calories (per 100g)?	Paprika Yellow	Paprika Red	Paprika Green	28	33	20
16	Which food has more calories (per 100g)?	Rhubarb	Radicchio	Peperoni	14	13	20
16	Which food has more calories (per 100g)?	Zucchini	Spinach	Pak Choi	18	15	16
16	Which food has more calories (per 100g)?	Leek	Broccoli	Red cabbage	24	26	22
16	Which food has more calories (per 100g)?	Wild garlic	Eggplant	Artichoke	19	17	22
17	Which country is older?	Albania	Finland	Hungary	1912	1917	1918
17	Which country is older?	New Zealand	Norway	Panama	1907	1905	1903
17	Which country is older?	Ghana	Niger	Togo	1957	1958	1960
17	Which country is older?	Tanzania	Ruanda	Mali	1964	1962	1960
17	Which country is older?	Brazil	Uruguay	Costa Rica	1822	1825	1821
18	Which country has more internet users (2015)?	Austria	Germany	UK	83.1	88.4	91.6
18	Which country has more internet users (2015)?	Luxemburg	Netherlands	Denmark	94.7	95.5	96
18	Which country has more internet users (2015)?	Portugal	Italy	Greece	67.9	62	63.2
18	Which country has more internet users (2015)?	Myanmar	Laos	Nepal	12.6	14.3	18.1
18	Which country has more internet users (2015)?	Jamaica	Peru	Panama	53.6	52.6	52
19	Which country has more analphabets (relative)?	Guinea	Niger	Burkina Faso	74.7	84.5	71.3
19	Which country has more analphabets (relative)?	Mali	Chad	Ethiopia	66.4	62.7	61
19	Which country has more analphabets (relative)?	Liberia	Haiti	Sierra Leone	57.1	51.3	55.5
19	Which country has more analphabets (relative)?	Pakistan	Bhutan	Senegal	45.3	47.2	47.9
19	Which country has more analphabets (relative)?	Nigeria	Mozambique	Gambia	48.9	49.4	48

Sources: https://de.wikipedia.org, http://de.statista.com, http://www.lebensmittel-tabelle.de, http://www.welt-in-zahlen.de, http://www.internetworldstats.com

Table 16: List of pre-round training icons

Set	Theme	Icon 1	Icon 2
1	Dots	1 black dot	3 black dots
2	Lines	2 horizontal lines	4 horizontal lines
3	Arrows vertical	up	down
4	Shapes 1	circle	square
5	Operators	plus	minus
6	Balls	soccer ball	basket ball
7	Pets	cat	dog
8	Gathering	sitting	standing
9	Travel	lake	mountains
10	Evening activity	board game	listening to music
11	Food	pizza	pasta
12	Exercising	dancing	running
13	Winter sports	skiing	snowboarding
14	Summer sports	swimming	cycling
15	Seasons	summer	winter
16	Story	book	movie
17	News	newspaper	smartphone

C.3 Instructions of Experiment 1

Below we present the translated instructions (originally in German) for the *S1 Reward* treatment of Experiment 1. The other treatments deviate from the instructions presented in the following ways:

- In each session, we conducted both the *Facts* and *Taste* domains, and we varied the order. The instructions for the domain that was conducted first were presented in detail, and participants received shortened instructions for the domain conducted second. In what follows, we show the extensive instructions for the *Taste* domain and the short form instructions for the *Facts* domain.
- There was no Stage 3 in the *Control* treatments.
- In the *Punishment* treatments, we talk about a deduction (instead of a bonus) of 10 points.
- The text in blue applies to the *S1* and *S2* treatments and is omitted in the *S0* treatments. The instructions of the *S1* and *S2* treatments were identical, but an additional page of instructions was shown before the experiment in *S2* as provided in Subsection C.4.

Instructions

Please keep quiet in your cubicle and do not communicate with others during the experiment. Anyone who intentionally violates this rule will be asked to leave the experiment without payment.

If you have any questions, please raise your hand and wait for an experimenter to come to you.

The incomes will be calculated in points. At the end of the experiment, the total amount of points you have earned will be converted into euros according to the following rate:

$$1 \text{ point} = 1 \text{ euro}$$

You will receive your total income in cash at the end of the experiment.

Please read the instructions carefully. Once everyone has finished reading the instructions, you will answer some comprehension questions. Then you will make your decisions in the experiment. Your decisions will be treated anonymously.

General procedure

This experiment consists of two parts, each comprising three stages. In each stage, you will make several decisions. Your total income is the sum of your income from both parts.

At the beginning of the first part, you will be randomly divided into groups of three. At the beginning of the second part, you will again be divided into groups of three.

Below you will find the instructions for Part 1. You will receive the instructions for Part 2 when Part 1 is completed.

Your decisions in the first part do not affect your income in the second part.

Which postcard do you choose?

Overview

In this part, you choose between two art postcards. The paintings of the two cards are displayed on the screen and you choose which of the two motifs you prefer.

After all group members have made their decisions, the motifs selected by the three persons are shown to an evaluator. Based on the selected motifs, the evaluator marks one person of your group, who may then receive a bonus. At the end, an evaluator whose decision is relevant for the bonus in your group is randomly selected. The more other evaluators select the same person, the higher the payout of an evaluator.

These decisions are made for several pairs of postcards.

At the end of the experiment, one decision situation will be randomly selected for your group. You will receive your preferred motif from this situation as a real postcard. For each group, a different pair of postcards will be randomly drawn, from which the group members will receive their preferred card. Thus, only members of your group can potentially receive the same postcard as you at the end of the experiment.

You will receive 10 points for participating in this part. If you are the person marked in the selected decision situation, 10 more points will be added to your account. In addition you decide as an evaluator for other groups. The more similarities you have in your decisions with other evaluators, the higher your payout as an evaluator will be.

In this part, you will go through three stages, which are described in more detail below.

Stage 1

You will see two postcard motifs on the screen, as shown in Figure 1. You decide which postcard you prefer to have by clicking on the corresponding motif.

After each decision, we will ask you to indicate how strong your preference is for the motif you have selected. To do so, once you have made your decision, a bar will appear below the motifs as shown in Figure 1.

You will make these decisions sequentially for 20 pairs of postcards. The members of a group are sometimes given different pairs to choose from.

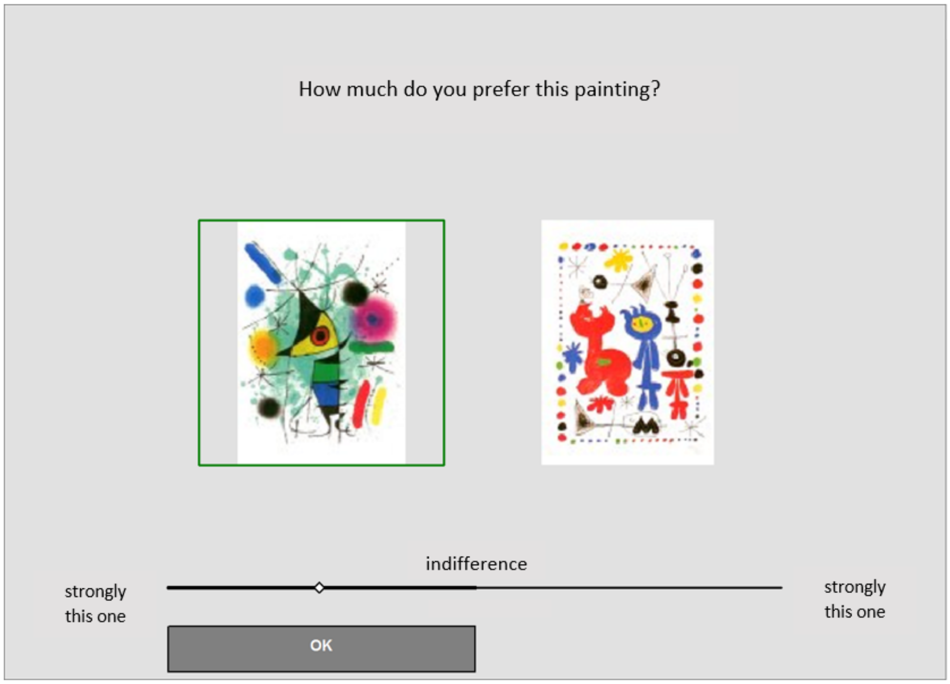


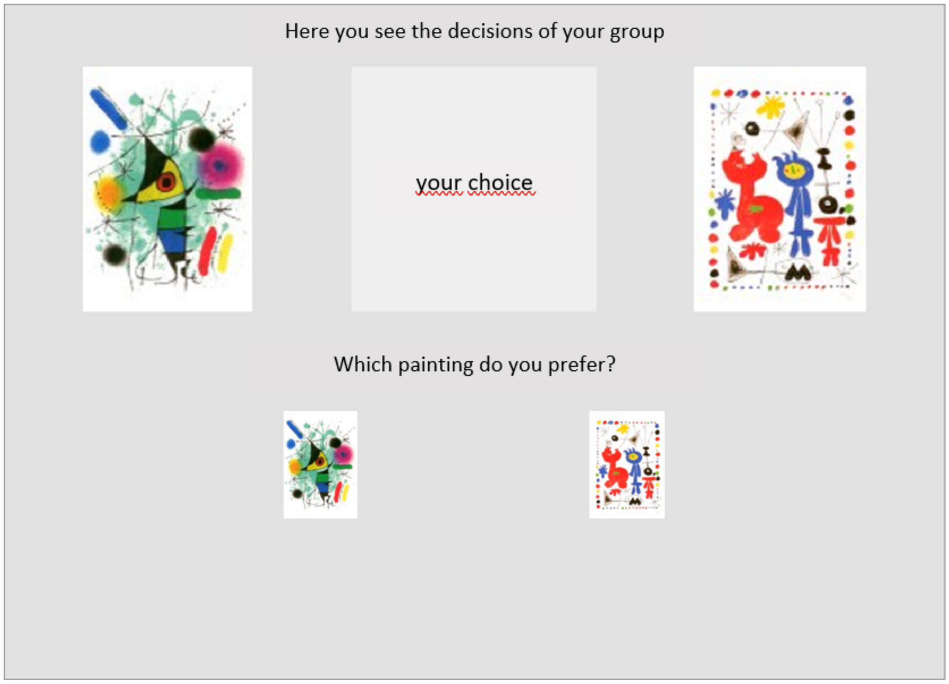
Figure 1

Stage 2

In this stage, you also choose one of two postcard motifs. The other two members of your group have already gone through the decision-making situations in their Stage 1 that you face in Stage 2. Before each decision, you will see how your group members have decided on the respective pair of postcards (upper part in Figure 2a). Again, you select a motif and indicate how strongly you prefer that motif (lower part in Figure 2b).

You make this decision in a sequence for 10 pairs of postcards.

Along with your decision, in the top row, you will see the 3 postcards that were selected by your group for the respective pair of postcards (upper part in Figure 2b). These 3 postcards are then sent to the evaluators in Stage 3, where the order of the 3 postcards on the evaluators' screens is random and can be different for each evaluator.

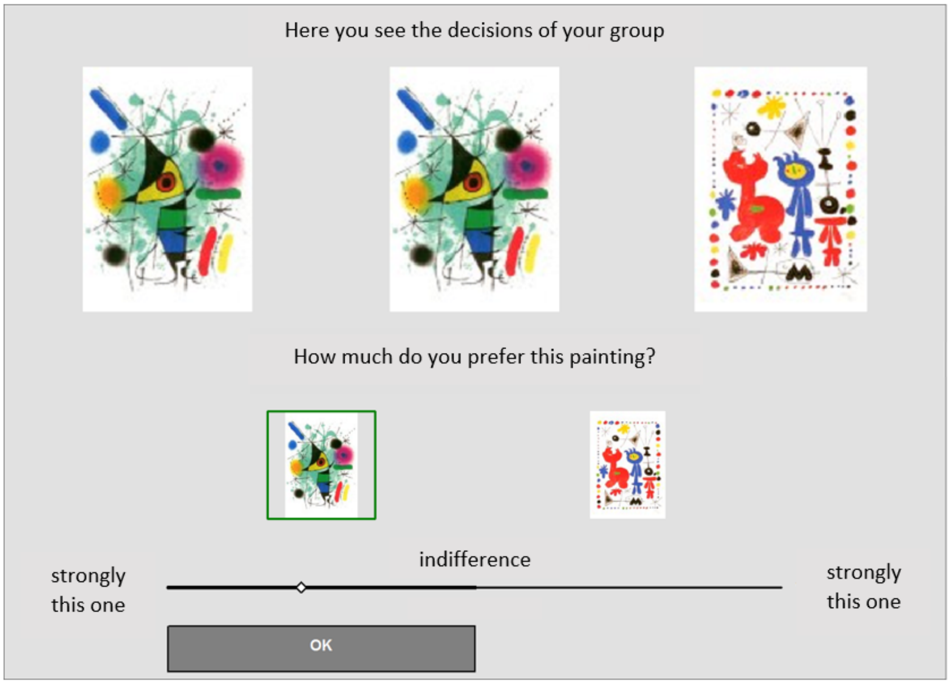


Here you see the decisions of your group

your choice

Which painting do you prefer?

Figure 2a



Here you see the decisions of your group

How much do you prefer this painting?

Figure 2b

Stage 3

For a given decision situation, the 3 selected motifs of your group will be sent to members of other groups for evaluation. Based on the selected motifs, each evaluator is asked to mark a person in your group, who may then receive a bonus of 10 points. An evaluators' payoff is higher the more of the other evaluators mark the same person as she/he does.

You will also decide as an evaluator. For a given decision situation, you will see how the three members of another group have decided, and on the basis of the selected motifs, you will mark who should receive the bonus. The more of the other evaluators make the same decision as you do, the higher your payoff as an evaluator.

At the time of your decision as the evaluator, however, you do not yet know which decision situation in a group will be randomly selected for payment and how the three group members actually decided in this situation. You therefore indicate who should receive the bonus for several possible constellations (see Figure 3).

The positions where you see the preferred postcards of the three group members are determined randomly. Thus, the selected motifs of a person sometimes appear on the left, sometimes in the middle and sometimes on the right, and the positions are shuffled for each decision and each evaluator.

At the end, a random draw is made to determine which decision situation and which evaluator will be relevant for your group. One member of each group will receive the bonus.

Example: Below you see various situations that may arise in a group when choosing between two postcards. In Figure 3a, all group members have chosen the same postcard. In Figure 3b, two people chose one postcard and one person has chosen another. As the evaluator, you will mark who should receive the bonus. To do so, click on the corresponding motif. Your selection will be highlighted by a green frame.

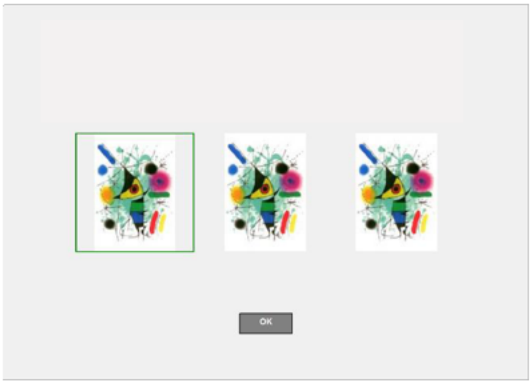


Figure 3a

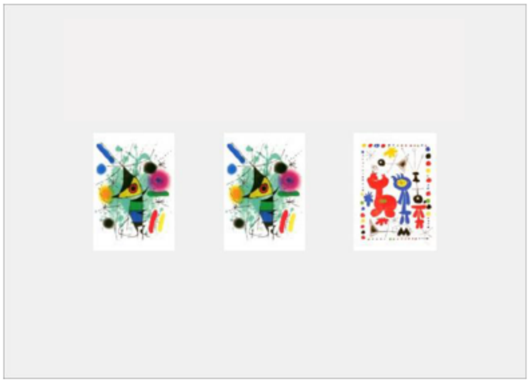


Figure 3b

Ultimately, the constellation that actually occurred in the group randomly assigned to you always applies. For example, if the group members have decided as shown in Figure 3a and have all selected the same motif, the person you marked for this constellation may receive the bonus.

Like all decisions, the evaluators' decisions are also mutually anonymous. Neither the selected person nor the evaluator will ever know the identity of the other person.

Other evaluators also decide who should receive the bonus for the same situations as you. All evaluators receive additional payments for their decisions. These payments are higher the more matches you have with other evaluators.

Concretely, you (and all other evaluators) receive 0.02 points per 10% matches for each situation. So if in a situation 10% of the other evaluators have marked the same participant as you, you will receive 0.02 points; if you match half (50%) of the others, you will receive 0.1 points; and if you match all of the other evaluators (100%), you will receive 0.2 points. According to this principle, your payoff is calculated and added up for each evaluation situation.

Please note that the displayed order of participants is random and may be different for each evaluator.

End

Finally, one decision situation per group will be selected at random. The motifs of one group are not used for another group. Therefore, it is only possible for the members of your group to receive the same postcard to take home.

You will then learn which postcard you will receive based on your decision in the randomly drawn decision situation. You will also be informed whether you were the person marked in this decision situation and thus receive a bonus.

For each of your decisions as an evaluator, you will learn to what extent your choice matches with other evaluators and what payment you will receive for this.

You will receive your postcard at the end of the experiment together with the payment.

If you have any questions, please raise your hand at any time.

Once you have read and understood the instructions, click on the “Experiment” button at the top right and then on the “Ready” button.

You can also access the instructions during the experiment. Please make sure that you do not miss out when the experiment continues.

Which answer do you choose?

For this part, you will be divided into a new group of three. The members of your new group were in three different groups in the first part.

All the procedures in this part are the same as in the previous part - with one difference: you do not decide between art postcards, but between two answers to a facts question. The question and the two answers are displayed on the screen and you choose one of the answers.

After all group members have made their decisions, the answers selected by the three individuals are shown to evaluators from other groups. Based on the answers selected, each evaluator marks one person in your group who may then receive a bonus. The order of the 3 answers on an evaluator's screen is again random and reshuffled for each decision situation and for each evaluator. As an evaluator, you also decide for other groups. All evaluators again receive an additional payment, which is higher the more often your selection matches with other evaluators.

These decisions are made for several facts questions.

At the end of the experiment, one decision situation will be randomly selected for your group. You will then be informed whether your answer in this situation was correct. A different facts question will be drawn at random for each group.

You will receive 10 points for participating in this part. If you are the person marked by the randomly selected evaluator in the randomly selected decision situation, additional 10 points will be added to your account. For your decisions as an evaluator, you will again receive 0.02 points per 10% matches with other evaluators.

This part also consists of the three stages that you have already completed in the previous part.

Once you have read and understood the instructions, click on the “Experiment” button at the top right and then on the “Ready” button.

C.4 Instructions of salience training rounds (S2 treatments)

Pre-rounds

In these pre-rounds of the experiment, you will be shown three images each on the screen, with the same motif appearing multiple times. You will mark one of these images. The more of the other participants have marked the same image as you, the higher your payoff from the pre-rounds.

You will make these decisions for multiple motifs. All decisions remain anonymous.

Example: Below you see two different situations that can occur. In Figure 1a, you see the same image three times. In Figure 1b, you see the same image twice and a different image once. In each setting, you will mark one of the three pictures by clicking on it. The other participants also mark one of the three pictures in the same settings.

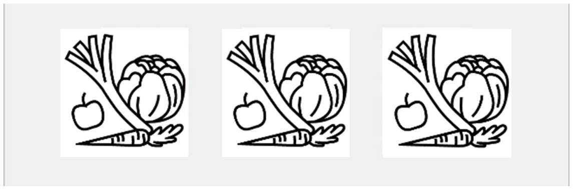


Figure 1a

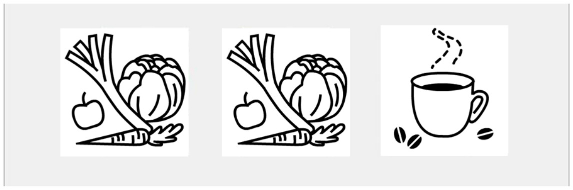


Figure 1b

Note that the displayed order of the images is random and may be different for each participant. For each decision and each participant, the positions are reshuffled. Thus, the same image appears sometimes on the left, sometimes in the middle, and sometimes on the right. The image that appears in the center for you may appear on the left or on the right for other participants. This applies to situations like in Fig. 1a as well as to situations like in Fig. 1b.

Once everyone has made their decisions in a given round, you will learn how the other participants have decided. All participants receive the payoffs corresponding to their decisions. These payoffs (1 point = 1 euro) are higher the more matches you have with other participants.

Concretely, for each situation, you (and everyone else) will receive 0.02 points per 10% matches. So, if in a given situation 10% of the other participants have marked the same picture as you, you will get 0.02 points; if you match half (50%) of the others, you will get 0.1 points; and if you match all the other participants (100%), you will get 0.2 points. According to this principle, for each round your payout is calculated and added up.

If you have any questions, please raise your hand.

Once you have read and understood the instructions, click on the ”Experiment” button at the top right and then on the ”Ready” button.

You can also access the instructions during the experiment. Please make sure you don't miss out when the experiment continues.

C.5 Instructions of Experiment 2

Below we present the translated instructions (originally in German) for the *S1 Reward* treatment of Experiment 2. The other treatments deviate from the instructions presented in the following ways:

- In the *Punishment* treatments, we talk about a deduction (instead of a bonus) of 2 points, and the flat payment was 12 points.
- The text in blue applies to the *S1* treatment and is omitted in the *S0* treatments. There was one evaluator per group in the *S0* treatments.

Each session consisted of three parts: the main treatments (Part 1, as shown below); the Krupka-Weber tasks and color ratings (Part 2); and post-experimental questionnaires (Part 3).

Instructions

Please read the instructions carefully. If you have any questions, please raise your hand and wait for an experimenter to come to you.

Please keep quiet in your cubicle and do not communicate with others during the experiment. Your cell phones should now be switched off. If you are carrying a device that is switched on, please switch it off immediately and place it in the holder provided. Anyone who intentionally violates this rule will be asked to leave the experiment without payment.

Your incomes will be calculated in points. At the end of the experiment, the total amount of points you have earned will be converted into euros according to the following rate:

$$1 \text{ point} = 1 \text{ euro}$$

You will receive your total income in cash at the end of the experiment.

All your decisions as well as your payoff will be treated anonymously.

The experiment consists of three parts. On the following pages you will find the instructions for the first part. You will receive the instructions for the second and third parts once the first part has been completed. Your decisions in the first part do not affect on your income in the following parts.

Part 1

Division into groups

Before the experiment starts, you will be divided into groups of 6 people. In Part 1, you will only interact with participants from your own group. There are two roles, *designers* and *evaluators*. Each group consists of 4 designers and 2 evaluators. You will be informed of your role before Part 1 begins. The assigned roles remain the same throughout the experiment.

In the beginning, each designer receives an endowment of 12 points, and each evaluator also receives an endowment of 12 points.

General procedure

Part 1 consists of 8 rounds. Each round follows the same procedure and consists of four phases: Design phase, publication phase, evaluation phase and feedback phase. In the *design phase*, each designer generates several colors. In the *publication* phase, each designer publishes one color which will be shown to the evaluators. In the *evaluation phase*, each evaluator then selects a designer on the basis of the four published colors. If both evaluators choose the color of the same designer, they receive an additional payment. Moreover, one evaluator will be randomly drawn whose decision determines which of the designers receives a bonus of 2 points. In the *feedback* phase, the designers learn who receives the bonus.

I Design phase

In the design phase, each designer generates colors by mixing them. In each round, all designers have 2 minutes to do so. During this time, the evaluators may also generate colors to pass the time. Figure 1 explains the screen on which the colors are generated. The screen consists of 3 areas: workspace, selection area and history.

Workspace: New colors are generated in the lower left workspace. To do so, hold down the left mouse button and drag a color from the color palette or the clipboard into one of the fields of the color bar. The two colors stored in the color bar are mixed together and the result appears directly below as a mixed color. To further process mixed colors further, they can first be dragged to the clipboard by holding down the left mouse button and then used again for mixing.

Selection area: In the selection area at the bottom right, the designers can store colors that they are considering for publication. Using the arrow keys, the current mixed color can be loaded into one of the three memories (△), or a color can be loaded from a memory for editing as a mixed color (▽). The double arrow can be used to exchange the mixed color with the color in a memory.

History: In the history, starting with round 2, the designers will see all colors published by the designers in their group in the previous round. Their own color is marked by a symbol, and the color that had been selected for the bonus is outlined in grey.

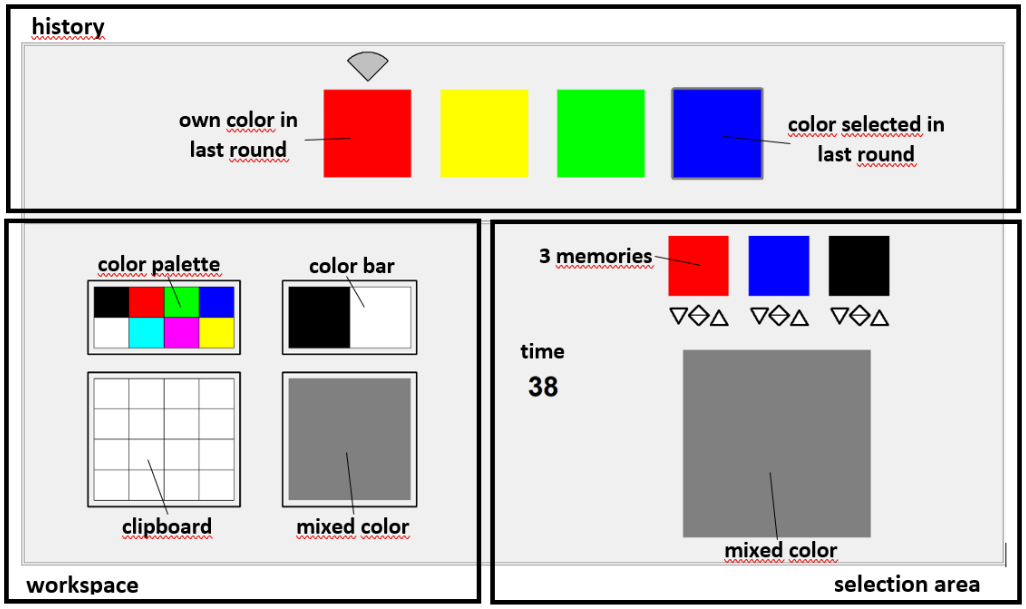


Figure 1: Screen for color generation

II Publication phase

One color from each designer will be published in each round. Only colors that are in the designer's selection area at the end of the design phase can be published. These are the colors in the three memories of the selection area as well as the current mixed color. From these four colors, each designer first makes a *preselection* and then a *conditional selection*.

Preselection: When the time of the design phase has expired, each designer first makes a preselection. To do so, they select one of the four colors in their selection area. The colors preselected by the designers are then temporarily displayed in the history for all other designers of their own group. The evaluators will not see the preselected colors.

Conditional selection: In the conditional selection that follows, each designer can adjust their decision based on the displayed results of the preselection. To do this, they again select one of their four colors.

Submission: The preselected colors of three designers are now submitted to the evaluators. However, in each round, one designer will be randomly chosen whose conditional selection referring to the three others' submitted pre-selected colors will be submitted.

All colors not selected by the designers remain private. This means that no other participant will see them at any time during the experiment.

III Evaluation Phase

The four colors submitted by the designers in a group are now displayed to the evaluators (Figure 2a). The arrangement of the colors is determined randomly in each round and for each evaluator. The position of a designer's colors therefore changes both across the rounds and across the evaluators. Therefore, a position is uninformative of a designer's previous publications. Moreover, it is likely that for the two evaluators, the same position will show different colors.

Each evaluator now selects by mouse-click one of the colors. The evaluators do not know whether a color is from the preselection or the conditional selection. The two evaluators receive an additional payment of 2 points if both have selected color of the same designer. In the example in Figure 2b, the relevant evaluator has chosen the yellow color (indicated by the grey border). Only if the irrelevant evaluator has also chosen yellow, both evaluators will receive an additional 2 points. If the evaluators have chosen different colors, they do not receive an additional payment. This does not affect the designer's payoff.

Before the first round begins, a random draw determines one of the two evaluators whose decision will be relevant for the designers. This relevant evaluator is the same person in all rounds. The other evaluator is irrelevant for the designers. However, the evaluators themselves do not know which of the two is the relevant evaluator. The designer of the color selected by the relevant evaluator receives a bonus of 2 points. This does not affect the evaluators' payoffs.

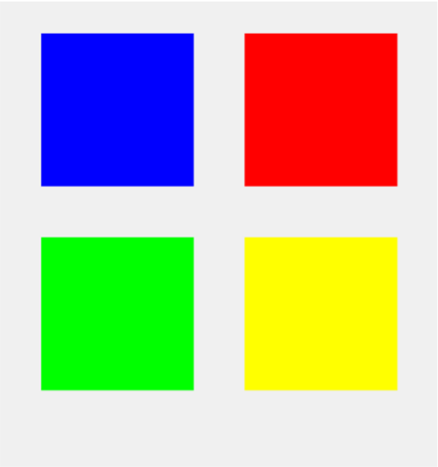


Figure 2a: evaluator selection screen

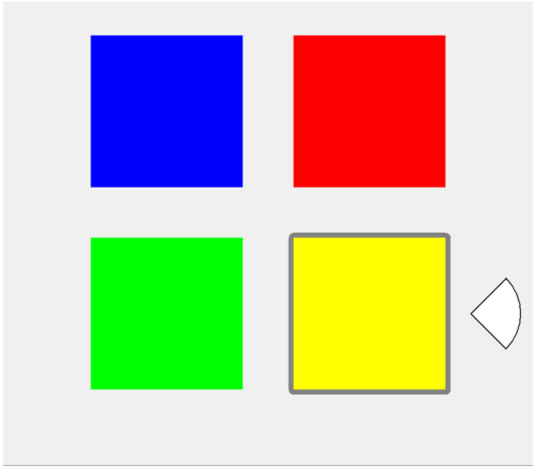


Figure 2b: designer feedback screen

IV Feedback phase

Once the evaluators have made their decisions, the designers will see the decision of the relevant evaluator (Figure 2b, grey border). The designers do not learn about the decision of the irrelevant evaluator. A designer's own color is marked by a white symbol. This way, designers can see whether they have received the bonus in case several identical colors were submitted.

At the end of a round, the evaluators do not yet know the other evaluator's choice. Only at the end of the experiment do they find out how often both have chosen the same designer and what payment they receive for that.

If you have any questions, please raise your hand. Once you have no more questions and are ready for Part 1, please click on "Experiment" in the upper right corner and then on "Inform".