No. 122

THURGAU INSTITUTE OF ECONOMICS
at the University of Konstanz

Fabian Dvorak
Urs Fischbacher
Katrin Schmelz

Incentives for Conformity and Anticonformity

Research Paper Series
Thurgau Institute of Economics
and Department of Economics
at the University of Konstanz

Incentives for Conformity and Anticonformity*

Fabian Dvorak
University of Konstanz
fabian.dvorak@uni.kn

Urs Fischbacher
University of Konstanz
urs.fischbacher@uni.kn

Katrin Schmelz
University of Konstanz
katrin.schmelz@uni.kn

December 24, 2020

The latest version is available here.

Abstract

We study how social evaluation affects conformity and anticonformity in theory and in an experiment. In theory, we show that negative social evaluation, i.e., potential punishment, creates incentives for conformity. Positive social evaluation, i.e., potential reward, creates incentives for anticonformity. In a laboratory experiment, we investigate the effect of these incentives in three domains: judgments in the knowledge domain, subjective arts preferences, and decisions in a creativity-related task. We rely on a new design in which we compare choices under social influence with predictions based on choices without social influence using transitivity. The experimental results confirm the theoretical predictions.

Keywords: conformity, creativity, social learning, institutions

JEL Classification: C92, D83, D91

*We would like to thank Charles Efferson, Armin Falk, Simon Gächter, Ronald Hübner, Willemien Kets, Simeon Shudy, and Egon Tripoldi for helpful comments and suggestions, and Ro'i Zultan for pointing out the quote in the introduction. All remaining errors are our own.

1 Introduction

What do you recommend wearing to a job interview? 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. (Rubinstein, 2013, pp. 195-196)

It is quite uncommon to make recommendations against conformity, in particular for a high stake situation such as a job interview. But there is a good argument against conformity that Rubinstein comes up with in this quote. If the applicant decides against conformity, she is special, she stands out, which might increase the probability to be selected. Thus, if being selected is good, conformity might not be the promising way to go. The situation looks quite different if being selected is bad. For example, if the company plans layoffs, conformity might be protective while standing out could provide a reason for being selected.

Our paper builds on the common understanding that one should conform in a high stakes situation like a job interview but turns the argument around. There is strong evidence in the literature that people often tend to conform, and that there are strong incentives to do so. But the quote also shows that the intuition to conform may be wrong in a situation where one person is selected from a group to receive a benefit. Ever since the seminal studies of Asch (1952, 1956), conformity is generally considered to be a strong behavioral motive and abundant evidence for conformity has been produced (see Cialdini and Goldstein, 2004, for a review), both in the lab and in the field. For example, when being informed that most others will do so, people are more likely to pay taxes (Bobek et al., 2007; Coleman, 2007), save energy (Allcott, 2011; Allcott and Rogers, 2014; Nolan et al., 2008; Schultz et al., 2007) or donate to a charity (Alpizar et al., 2008; Smith et al., 2015).

The pervasiveness of social evaluation among humans has been put forward to explain the ubiquity of conformity in general. Social evaluation is the spontaneous cognitive process of identifying friend and foe on the basis of behavior or physical appearance (Ambady and Rosenthal, 1992; Cunningham et al., 2004; Winter and Uleman, 1986). Humans constantly evaluate the actions and intentions of others in order to make accurate decisions about who deserves punishment and who deserves reward, which is crucial for the enforcement of social norms (Fehr et al., 2002; Fehr and Fischbacher, 2004; Gürerk et al., 2006; Guzmán et al., 2007; Henrich et al., 2006). The notion that conformity helps individuals to attract reward and avoid punishment when being evaluated by peers has a long history in social psychology (Festinger, 1953; Kelman, 1961; Allen, 1965).

In this paper, we investigate how the social evaluation of choices creates incentives for strategic conformity and anticonformity. In our theoretical model, an evaluator assesses several individuals based on their choices. The task of the evaluator is to select one individual. We distinguish two environments, one in which the selection leads to a reward and one in which it leads to a punishment. We assume that the allocation of reward and punishment is determined by the preferences of the evaluator or based on the salience of the choice. If the evaluation is based on taste, the evaluator allocates reward to those with similar preferences and punishment to those with different preferences. If the evaluation is based on salience, the evaluator selects choices that stand out independent of the consequence of the selection.

The model predicts that the effect of social evaluation on conformity and anticonformity hinges on its consequences. In the salience based evaluation, the incentives are obvious: conformity in the punishment environment and anticonformity in the reward environment. The situation is similar if the social evaluation is based on the evaluator's preference, i.e., if she selects someone with a similar preference for reward and someone with a different preference for punishment. In this case, it is optimal to blend in the punishment environment. In the reward environment, anticonformity is optimal if the decision is difficult and if the minority is small enough. Hence, the model predicts that social evaluation induces strategic conformity and anticonformity. Conformity and anticonformity are strategic in the sense that they are rational responses to social influence in order to avoid punishment or attract reward.

Experiments in the spirit of Asch have shown that participants make "surprising" choices more often when they observe others to make such "surprising" choices. This evidence that people deviate from their "real" choice under social influence is what we consider as *conformity* in this paper. The opposite of this behavior, *anticonformity*, i.e., deviating from one's choice under social influence by *mis*matching the social information, is far less studied (Ariely and Levav, 2000; Fromkin, 1970; Imhoff and Erb, 2009; Lynn and Harris, 1997) and evidence is often anecdotal. In particular, in experiments similar to those by Asch, there is no room for anticonformity. Choosing the "non-surprising" option is not evidence for anticonformity but evidence for *independence*, which refers to behavior unaffected by social influence (Argyle, 1957; Crutchfield, 1962; Willis, 1963; Willis and Levine, 1976). Relying on surprising choices also raises the problem that these choices are rare by definition. Thus, if no deception is used, much data is necessary in order to also get such rare choices as social information. Further, in these experimental settings, anticonformity is difficult to detect. Distinguishing between anticonformity and independence requires to either create a situation in which anticonformity is sufficiently frequent to be detected on the aggregate, or to know on the individual level what people would do with and without social information. We will present an approach that allows identifying conformity and anticonformity on the individual level.

To put the theoretical predictions to a test, we conduct laboratory experiments. We study the effect of social evaluation on conformity and anticonformity in three domains: judgments in the knowledge domain, arts preferences, and decisions in a creativity-related task. We use two settings. In experiment one, participants make binary choices on judgments concerning difficult knowledge questions and on preferences over paintings. In experiment two, we investigate how choices evolve in a creativity-related task where participants generate and publish colors. To expose participants to social information, we form groups and inform participants about the decisions of the other group members in a given decision situation. To implement social evaluation, an external evaluator, i.e., an evaluator who is not part of the group, selects one group member based on the choices of the group.

In the punishment treatments, the selected group member loses a fixed amount of experimental currency. In the reward treatments, the selected group member gains a fixed amount of experimental currency. To assess non-strategic (baseline) conformity, we also conduct a control treatment in experiment one in which participants' choices are disclosed to the other group members but not evaluated. Moreover, to create incentives for the evaluators to use salience as a selection criterion, we implement variants of the punishment and reward treatments in which several evaluators face the same set of choices and are incentivized to coordinate their selection.

Conformity and anticonformity are defined as a change in one's behavior due to social information. Identifying these phenomena requires to know an individual's preference in the absence of social information. Giving participants the same choice twice might trigger concerns for consistency. For this reason, we use predictions based on transitivity and define deviations from these predictions as conformity or anticonformity. This approach also allows us to distinguish conformity and anticonformity from independence.

Our central result is the following: Across all three experimental tasks, social evaluation increases conformity if its consequence is punishment, and anticonformity if the consequence is reward. The higher level of conformity in the treatments with punishment indicates strategic conformity to avoid being punished. The higher level of anticonformity in the treatments with rewards indicates strategic anticonformity to attract reward. Further, conformity and anticonformity vary across domains. We find the highest level of conformity when making judgments about difficult knowledge questions, which indicates social learning (Banerjee, 1992; Bikhchandani et al., 1992; Lee, 1993, 1998; Smith and Sorensen, 2000; Vives, 1997; Banerjee and Fudenberg, 2004; Anderson and Holt, 1997). In this task, participants know that an objectively correct answer exists and can thus update their belief about what is correct after observing others' choices. However, we also find conformity in preferences over paintings. This is in line with the insight that preferences are, to some extent, influenced by observations of others' behavior (Bandura and Walters, 1977), and with the theory and experiments on frequency dependent learning (Boyd and Richerson, 1982, 1985; McElreath et al., 2008; Efferson et al., 2008; Bernheim and Exley, 2015).

We also find meaningful individual variation in the responses to social information that is related to individual traits. Strong anticonformity is rare, and only displayed by a minority of participants in the reward treatments. Concerning the social evaluation, our results suggest that the main underlying mechanism is taste-based allocation of punishment and reward. Salience plays a minor role for the social evaluation of participants' choices even in the treatments with coordination incentives for the evaluators.

Related literature

Our study contributes to different research directions around our question on how social evaluation affects conformity and anticonformity. Most closely related to our study is Griskevicius et al. (2006), who use priming to investigate the effect of a self-protection versus a mate-attraction mindset on conformity and anticonformity in a picture selection task. The authors find that the self-protection mindset induces conformity and the mate-attraction mindset anticonformity (but only in men). From the perspective of our study, these results could be explained by the fact that the priming of self-protection induces the anticipation of punishment while the priming of mate-attraction induces the anticipation of reward. Similar effects of expected punishment and reward are found in behavioral ecology. The selfish-herd hypothesis suggests that the evolutionary benefit of forming groups is a reduced risk to die for the individual in the collective as the risk that an individual of the group is taken by a predator is distributed over all individuals (Hamilton, 1971; Vine, 1971). Studies in fish and crabs have shown 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 et al., 2017).

Previous research on how social evaluation creates social influence is restricted to the impact on conformity (Amabile et al., 1990; Shalley and Perry-Smith, 2001; Sakha and Grohmann, 2016; Bernheim, 1994), and does not consider anticonformity. What also sets our study apart is that we focus on the case where the selection of an individual in a group creates the incentive. This setting is frequent when the implementation of punishment and reward is costly for the evaluator or subject to institutional regulation. Another difference is that we focus on situations in which the evaluated individuals have incomplete information about the preferences and judgments of the evaluator. If the evaluated individuals know the preferences and judgments of the evaluator instead, they should align their behavior accordingly. Robin et al. (2014) show that social evaluation leads to ingratiation, and in this respect to conformity, if the preferences of the evaluator are known to the evaluated individuals.

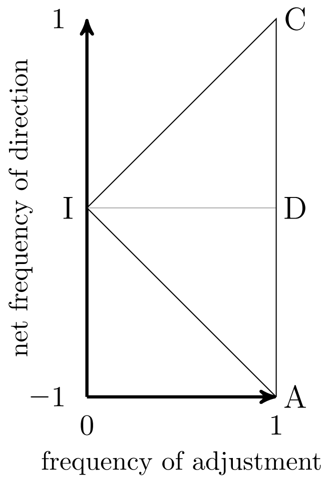
Similar to previous studies, we also find that individual differences in the response to social influence can be related to individual traits (Ariely and Levav, 2000; Fromkin, 1970; Imhoff and Erb, 2009; Lynn and Harris, 1997). This evidence is in line with the idea of preferences for conformity or anticonformity (Argyle, 1957; Brehm, 1966; Corazzini and Greiner, 2007; Fatas et al., 2018; Goeree and Yariv, 2015; Jones, 1984; Jones and Linardi, 2014; Wright et al., 2009), which can explain cultural variation in the response to social influence (Bond and Smith, 1996; Cialdini et al., 1999; Kim and Markus, 1999; Yamagishi et al., 2008).

Conformity and anticonformity determine the level of diversity in social groups. Therefore our results also relate to the normative question of implementing an optimal degree of diversity in social groups, organizations or society (March, 1991; Kets and Sandroni, 2016). Institutions which promote innovation and progress in society often award benefits to single (or few) individuals in the context of a public competition for a prize. Such rewards for individuals assures a desirable variety of entries and promotes different solutions to existing problems.

2 Conceptual framework

Our conceptual framework of conformity and anticonformity is the Willis-Nail model of social response (Willis, 1965; Willis and Levine, 1976; Nail, 1986; Nail and Van Leeuwen, 1993). The Willis-Nail model is illustrated in Figure 1. The model space is an isosceles right triangle spanning over two orthogonal dimensions: the horizontal independence-dependence dimension, and the vertical conformity-anticonformity dimension. The three vertices of the triangle represent the three canonical responses to social influence: conformity (C), anticonformity (A), and independence (I). The conceptual framework covers a fourth response to social influence, dependence (D) which is not particularly interesting in itself but conceptually necessary to cleanly separate the other three responses.

Figure 1: The triangle model of social response



Notes: The isosceles triangle model of social response (Willis, 1963; Nail and Van Leeuwen, 1993). C = conformity, A = anticonformity, I = independence, D = dependence.

The possibility to cleanly separate conformity, anticonformity, and independence is an advantage of the Willis-Nail model over one-dimensional models of conformity and anticonformity (Allport, 1934; Sherif, 1935; Asch, 1951; Argyle, 1957; Worchel and Brehm, 1970). While some unidimensional models do not take independence into account, others define independence as the absence of conformity or anticonformity (see Nail et al., 2013, for a discussion). The first approach is problematic since independent behavior is misrepresented as either conformity or anticonformity. The second approach is also problematic as mixed observations of conformity and anticonformity are misrepresented as independent. The Willis-Nail model overcomes these problems by introducing dependence as socially dependent behavior which is neither conformity nor anticonformity.

Operationalization

Our operationalization of the two orthogonal model dimensions differs from previous studies (Willis, 1965; Willis and Levine, 1976; Nail and Van Leeuwen, 1993; Nyczka and Sznajd-Weron, 2013; Nyczka et al., 2018). To locate behavior in the two-dimensional space, we use an experimental technique suggested by Nail (1986). It requires that an individual takes the same decision twice, in the absence and in the presence of information about the behavior of others. In the following, we refer to the choice in the absence of social influence as the *nonsocial* choice. We refer to the choice in the presence of social influence as the *social* choice. If the two choices differ, this is interpreted as an adjustment of the social choice due to social influence.

The horizontal independence-dependence dimension of the model depicted in Figure 1 is operationalized as the relative frequency of adjustments observed in the population. The vertical conformity-anticonformity dimension is operationalized as the net frequency of the direction of the observed adjustments. This net frequency is defined as the relative frequency of adjustments which make the social choice more similar to the choices of others minus the relative frequency of adjustments which make the social choice more *dis*similar to the choices of others. The fact that the sum of these two frequencies yields the overall frequency of adjustments generates the triangular model space.

To fix ideas, assume we observe N pairs of nonsocial and social 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 nonsocial choice and the social choice of the ith pair of choices. We define the probability to observe an adjustment of the social choice *towards* the choices of others as the relative frequency of adjustments that decrease Δ:

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

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

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

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 social choice in both directions. 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 social 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 social choice *away from* the choices of others is possible. The corresponding equations are:

$$P_b(towards) = \frac{\Sigma_{i \in N^t} \mathbb{1}(\Delta_i^{ns} > \Delta_i^s)}{|N^t|} \tag{5}$$

and

$$P_b(away) = \frac{\Sigma_{i \in N^a} \mathbb{1}(\Delta_i^{ns} < \Delta_i^s)}{|N^a|} \tag{6}$$

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.

Four canonical responses to social influence

Based on this operationalization of the model dimensions, four canonical responses to social influence arise from specific patterns of adjustment. If social choices are never adjusted, the response to social influence is independence, located at the left vertex of the triangle. If social choices are adjusted if and only if this makes the choices more similar to others' choices, the response to social influence is conformity, located at the upper vertex of the triangle. If social choices are adjusted if and only if this makes the choices more *dis*similar to others' choices, the response to social influence is anticonformity, located at the lower vertex of the triangle. Finally, if choices are always adjusted, the response to social influence is dependence, located in between the conformity and anticonformity vertices.

3 Theory

We use a simple model with binary choices to investigate the effect of social evaluation. We consider two decision rules for the social evaluation: social evaluation based on salience and based on own taste.

Social evaluation based on salience means that the evaluator selects a person that stands out because of his or her choice. An individual stands out because of his or her chosen alternative is the one chosen by the minority of individuals. If there is no minority, the evaluator randomly selects one individual. No minority exists when all individuals choose the same alternative or when each alternative is chosen by the same number of individuals.

Social evaluation based on taste means that the evaluator rewards someone who has chosen the alternative she prefers (if this is possible), and that the evaluator punishes someone who has chosen the alternative she does *not* prefer. If all individuals choose the same alternative, one of the three individuals is randomly selected.

We assume that the consequences of punishment and reward in terms of utility outweigh the intrinsic utility of the alternatives and focus on symmetric subgame perfect equilibria of the model. In a nutshell, the theory shows that punishment creates a clear incentive for conformity under both decision rules. Reward incentivizes anticonformity for decisions under salience based evaluation. The allocation of reward based on own taste incentivizes anticonformity for difficult choices. The more evenly distributed the preferences of the individuals are over the two alternatives, the more difficult is a choice.

Model

Consider a binary choice between two alternatives. We assume that the players have a preference for one of the alternatives. We also assume that the preferences of the players are correlated in the following way. There is a generally preferred alternative. The probability equals $\frac{1}{2}$ for either alternative to be preferred. Each player prefers this alternative independently with a probability $p > \frac{1}{2}$. A high probability p means that the preferences in the population of players are quite aligned and the choice is rather easy. A low probability p means that the preferences in the population of players are quite mixed and the choice is rather difficult.

There are $N \geq 3$ players, A_1, ... A_{N-1}, and B. The A players (i.e. player A_1, ... A_{N-1}) make their choice simultaneously. Player B chooses after observing the choice of the A players. Player B is not informed about the identity of the A players. The evaluator assigns a reward or punishment to one of the N players without knowing who player B is.

We consider only symmetric equilibria with respect to the A players. Thus, the strategy of the A players is defined by the probability q_A to follow the own taste. We assume $q_A \geq \frac{1}{2}$, which excludes equilibria in which players coordinate by choosing the not preferred alternative. Since player B is not informed about the identity of the A players, B can only condition the choice on the number k of A players who decide according to B's taste. This means that B's strategy is determined by a function $\beta\{0, ..., N-1\} \to \{0, 1\}, k \mapsto c(k)$, where $\beta(k) = 0$ means that B follows his own taste, and $\beta(k) = 1$ means that B chooses against the own taste.

Evaluation based on salience

We first investigate salience based social evaluation. We derive the following propositions.

Proposition 1 (Salience under punishment)**.** *The A players follow their taste. If there is a strict majority, i.e. a majority of size $> \frac{(N-1)}{2}$, then B follows this majority, and decides according to the own taste otherwise.*

Proposition 2 (Salience under reward)**.** *The A players randomize with a probability of $\frac{1}{2}$. If there is a strict minority, i.e. a minority of size $< \frac{(N-1)}{2}$, then B follows this* *minority (which can be empty), and decides according to the own taste otherwise.*

Evaluation based on taste

We now consider social evaluation based on taste. We start with some terminology. Player B's decision is called *independent* if it coincides with the own taste. It is called *conformist* if it neglects the own taste and follows the choice of the majority. It is called *anticonformist* if it neglects the own taste and follows the choice of the minority.

Proposition 3 (Taste based punishment)**.** *Independent of the strategy of player B, the A players always follow their taste. B is always conformist; if there is a strict majority, i.e. a majority of size $> \frac{(N-1)}{2}$, then B follows this majority, and decides according to the own taste otherwise.*

For taste-based reward, the behavior of player B is more complex. The best-response of player B is either conformity, anticonformity or independence. The best-response of player B depends on k, p and the strategy of the A players defined by q_A. Due to the complexity of B's best-response, we do not provide a general solution for the behavior of the A players in equilibrium. However, for the case $N = 3$, which we study experimentally, we prove the following proposition.

Proposition 4 (Taste based reward)**.** *If $N = 3$, the two A players always follow their taste, independent of the strategy of player B. For $k = 0$, Player B's choice is conformist iff p is above some threshold $C(N, k) < 1$. Player B's choice is anticonformist iff $p > \frac{1}{2} + \frac{\sqrt{3}}{6}$. For $k = 2$, Player B's choice is anticonformist iff $p < \frac{1}{6}\left(3 + \sqrt{6\sqrt{3} - 9}\right)$. For $k = 1$, Player B's choice is independent.*

The proofs of the four propositions are in Appendix A.

4 Experimental design

In this section, we present our two experiments. In experiment one, participants make binary choices in the domains of arts preferences and judgments on knowledge questions. In experiment two, participants perform in an open task, which is designing colors. In the following, we first provide an overview of the tasks. We then show how we identify conformity, anticonformity and independence. Finally, we present our treatments, the procedural details and the implementation.

4.1 Experimental tasks

We conduct two laboratory experiments to study the effects of the social evaluation of choices on conformity and anticonformity across three different tasks. The three experimental tasks differ in several dimensions which enables us to investigate the validity of our theoretical predictions across different choice environments. One important difference between tasks concerns the existence of an objectively correct choice. Another important difference concerns the number of available alternatives in each choice.

In experiment one, we study judgments and preferences in two separate tasks. In the *questions* task, participants select one out of two answers to a difficult knowledge question. In the *paintings* task, participants choose one out of two art paintings with similar motives. Each participant of experiment one executes both tasks in a within-subject design. The order of the tasks is balanced across two experimental sessions. In one experimental session all participants start with the questions task and proceed with the paintings task. In the other experimental session, all participants start with the paintings task and proceed with the questions task. Lists of the questions and paintings used in experiment one can be found in Tables C.3-C.5 in Appendix C.

Experiment two contains the *colors* task. In this task, participants choose one out of four self-generated colors. At the beginning of the experiment, we supply every participant with eight colors: red, green, blue, yellow, magenta, cyan, black, and white. These eight colors represent the vertices of the three-dimensional rgb color space. Using a color generation interface, participants can generate new colors by additive mixing two colors. If a participant generates a new color, the participant can store and reuse the color to generate further colors. By repeatedly executing these steps, every color in the rgb color space can be approximated. A picture of the color generation interface can be found in in Appendix C and a movie on https://fdvorak.com/videos/creativity-task.mp4.

4.2 Measurement of the response to social influence

To measure the response to social influence, participants are assigned to groups of three in experiment one, and groups of four in experiment two. To generate social influence, we inform participants about the choices of the other group members prior to the own decision. Participants know that the choices of the other group members were made in the absence of social influence. We consider the choice a participant makes after being informed about the choices of the other group members as the social choice of the participant.

Across the two experiments, we use different experimental techniques to elicit a corresponding nonsocial choice for each social choice. Different techniques to elicit the nonsocial choices are required because the number of alternatives in each choice differs across the two experiments. The choices in the questions and paintings tasks of experiment one are binary. This enables us to predict a nonsocial choices for each social choice based on transitivity. In experiment two, four alternatives exist in each choice which requires direct elicitation of the nonsocial choices. Predicting the nonsocial choices reduces the problem that participants might want to appear consistent in their choices (as shown by Falk and Zimmermann, 2017). As we have no reason to believe that the prediction error differs across treatments, we think the prediction of nonsocial choices is desirable.

Prediction of nonsocial choices in experiment one

In experiment one, we predict the nonsocial choice based on two related choices of the same participant. The two related choices are elicited in the absence of social influence. The two related choices compare each of the two alternatives of the social choice to a common third alternative. In each of the two related choices, the participant additionally indicates the strength of her preference or judgment on a continuous scale (see Appendix C.1 for the decision screen).

The nonsocial choice is predicted by assuming transitivity of the related choices. If one of the alternatives of the social choice is preferred over the common third alternative and the other is not, the preferred alternative is the predicted nonsocial choice. If either both or none of the alternatives of the social choice are preferred over the common third alternative, we compare the self-reported strength of the preferences or judgments across the related choices to make a prediction. If both alternatives of the social choice are preferred over the common third alternative, the alternative that is preferred more strongly is the predicted nonsocial choice. If none of the two alternatives of the social choice is preferred over the common third alternative, the alternative that produces the lower preference strength for the common third alternative is the predicted nonsocial choice.

We also generate a prediction for the strength of the preference or judgment in the nonsocial choice by comparing the self-reported strength of the preference or judgment across the two related choices. We use the predicted strength of the preference or judgment to model participants' decisions to adjust the social choice. We expect that participants more frequently adjust the social choice if the predicted strength of the preference or judgment is low.

To calculate coordinates of the response to social influence for the binary choice data of experiment one, we focus on all social 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 which is positive if the social choice differs from the alternative chosen by both other group members and zero otherwise.

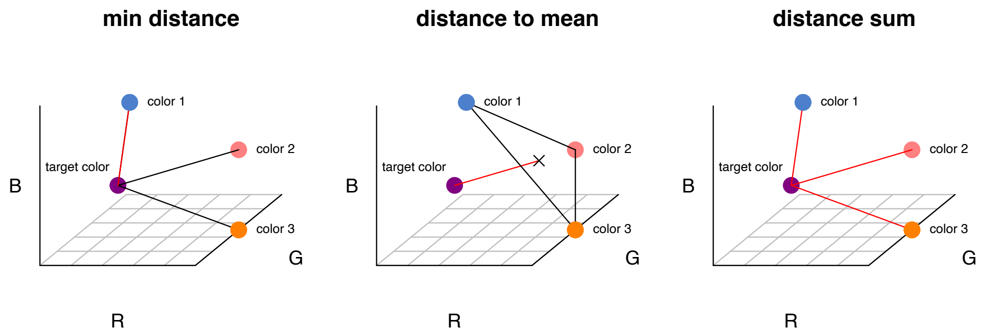
Elicitation of nonsocial choices in experiment two

For experiment two, we elicit participants' nonsocial choices which are subsequently transmitted to the other members of the group. Every group member makes her social choice after being informed about the nonsocial choices of the other group members. Participants know that the social choice of one randomly selected participant will be evaluated together with the nonsocial choices of the other group members.

To calculate coordinates of the response to social information for the multinomial choice data of experiment two, we focus on all situations in which the social choice can 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.

The identification of a distance variable which adequately measures the similarity of a target color to three other colors is not obvious. To explore the robustness of the experimental results with respect to the specification of the distance variable, we calculate the coordinates of the response to social influence based on three different distance variables and two different color metrics.

Figure 2: Distance variables for the similarity of one color to three other colors



Notes: Red lines in the three panels illustrate the three variables used to determine color similarity in experiment two 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.

Figure 2 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 2 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.

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.

While rgb color metric is the most common specification to measure color distance, it does not necessarily reflect perceived differences in colors. Therefore, we additionally calculate the values of the three variables illustrated in Figure 2 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).

In the statistical analyses, we mainly focus on the min distance variable which has been proposed as a measure for the cohesion of individuals in groups (Clark and Evans, 1954). We use the min distance variable in combination with the rgb metric as the latter is the most frequently used color difference metric. We explore the robustness of the experimental results of experiment two based on the five remaining combinations of the three distance variables and two color metrics.

4.3 Experimental treatments

In both experiments, we implement four treatments with social evaluation. In the treatments with social evaluation, the choices of a group of participants are evaluated by a participant who is not part of the group. For the purpose of evaluation, the choices of the group members are displayed on the screen of the evaluator (an example of the evaluation screen can be found in Appendix C.1). Each group member is represented on the screen by her choice. One of the choices displayed on the screen is a social choice. The other choices on the screen are made in the absence of social influence. The evaluator is not informed which of the displayed choices is the social choice.

In all four treatments with social evaluation, the task of the evaluator is to select one member of the group by clicking on one of the choices displayed on the screen. The four treatments with social evaluation differ with respect to the consequence of the selection for the selected participant, and with respect to the incentives of the evaluators. In two treatments, the consequence of being selected is punishment. In the other two treatments, the consequence of being selected is reward. In one punishment treatment one reward treatment, the evaluation decision is not incentivized. Evaluators are instructed to select what they find most appropriate in the given situation. In the other treatments, evaluators are incentivized to coordinate their evaluation decision with other evaluators. Evaluators are informed that their payoff increases if another evaluator selects the same participant.

The purpose of the treatments with coordination incentives is to induce an allocation of reward and punishment based on salience. We conducted the treatments with coordination incentives after the treatments without coordination incentives to investigate whether the allocation of punishment and reward based on salience would generate different results. In the treatments with coordination incentives, several participants evaluate the same choices and receive an additional payment for each other evaluator who selects the same participant. As it is not possible for the evaluators to communicate their evaluation decision to other evaluators, coordination of the evaluation decisions must be achieved by selecting the choice which is generally considered salient.

The four between-subjects treatments with social evaluation are summarized in Table 1.

In experiment one, we additionally conduct a control treatment without social evaluation. In the control treatment, choices are disclosed to the other group members

Table 1: Between-subjects treatments with social evaluation

	no coordination	coordination
punishment	Payoff of selected individual is decreased. Evaluation decisions not incentivized.	Payoff of selected individual is decreased. Evaluators incentivized to coordinate.
reward	Payoff of selected individual is increased. Evaluation decisions not incentivized.	Payoff of selected individual is increased. Evaluators incentivized to coordinate.

but not evaluated. We do not conduct a control treatment in experiment two.

4.4 Implementation

We conducted 26 experimental sessions between May 2016 and October 2018 with students of the University of Konstanz in Germany. Participants are recruited with ORSEE (Greiner, 2015). The experiments are conducted with z-Tree (Fischbacher, 2007). The combined data of both experiments contains the decisions of 745 students. Mean age of the participants is 21.6 years and 58.7 percent are female. Table 2 summarizes the number of sessions, participants, groups and the total number of choices in each experimental condition.

Procedures

In experiment one, participants first make 20 choices in the first task (arts preferences or knowledge questions, depending on the order of those treatments in their session) in the absence of social influence. These choices are subsequently used to predict 10 nonsocial choices for a given participant and to provide social information to group members in social choices of the same task. After the social choices are made, every participant of experiment one takes the role of an evaluator and evaluates the choices of another group. We use the strategy method (Selten, 1967) to elicit the evaluation decisions for every possible combination of social choices. Experiment one continues by repeating the same procedure for the second task. At the end of experiment one, participants receive feedback about the evaluation decisions and their payoff. We use the same materials and procedures in the control treatments of experiment one which do not feature an evaluation phase. We conducted one experimental session starting with the questions task and one starting with the paintings task for each treatment.

In experiment two, participants take the fixed role of a *designer* or an *evaluator* for eight rounds of the experiment. Groups consist of four designers and one evalua-

Table 2: Summary of treatment data

experiment	one										two			
task	questions					paintings					colors			
evaluation	reward		punish		no	reward		punish		no	reward		punish	
coordination	no	yes	no	yes	no	no	yes	no	yes	no	no	yes	no	yes
sessions	2	2	2	2	2	2	2	2	2	2	4	4	4	4
participants	54	51	54	51	60	54	51	54	51	60	115	120	120	120
groups	18	17	18	17	20	18	17	18	17	20	23	20	24	20
social choices	540	510	540	510	600	540	510	540	510	600	736	640	768	640
evaluations	6480	6120	6480	6120	-	6480	6120	6480	6120	-	184	320	192	320

Notes: The number of participants (groups) reflects the number of statistically independent observations in experiment one (two). In experiment one, all participants take the role of an evaluator. This together with the fact that evaluations were elicited using the strategy method leads to the relatively large number of evaluation decisions.

tor in the treatments without coordination incentives and four designers and two evaluators in the treatments with coordination incentives. The group composition remains fixed across rounds.

Designers have two minutes before every round to create new colors by mixing preexisting colors. After two minutes each designer makes the nonsocial choice by submitting one out of up to four different colors she created before the current or before any of the previous rounds. The nonsocial choices are subsequently disclosed to the other group members. Every designer makes the social choice. A random number selects the group member whose social choice is submitted along with the nonsocial choices of the other group members to the evaluator(s).

In the coordination treatments with two evaluators per group, the decision of one of the two evaluators is implemented. The evaluation decisions of the second evaluator do not affect the payoffs of the evaluated participants.

After each round, all participants receive feedback about the evaluation decision and the consequences for their payoff. The experiment ends after eight rounds.

Incentives

The choices in the paintings task of experiment one are incentivized by handing over one of the chosen paintings in the form of an art postcard at the end of the experiment. After the experiment, participants compare their answers in the questions task to the objectively correct answers. This is publicly announced before the experiment to trigger an intrinsic motivation to give correct answers. The choices in the colors task of experiment two are not incentivized.

In experiment one, participants receive a flat payment for the completion of each experimental task which varies over treatments. The flat payment is 30 Euro in the punishment and 20 Euro reward treatments, and 16 Euro in the control treatment without social evaluation. The flat payments are chosen such that average earnings are similar across treatments.

In the punishment and reward treatments of experiment one, we randomly select one evaluation decisions per group for the questions task and for the paintings task. The randomly selected evaluation decision is then implemented. In the punishment treatments, we deduct 10 Euros from the final payoff of the selected participant. In the reward treatments, we add 10 Euros to the final payoff of the selected participant. In the treatments with coordination incentives, the payoff of the evaluator 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. For example, if all evaluators select the same group member, the payoff of every evaluator is increased by $100 \times 0.002 = 0.2$ Euro.

In experiment two, participants also receive a flat payment in the beginning of the experiment. The flat payment is 20 Euros in the punishment and 12 Euro in the reward treatments. As in experiment one, the flat payments are chosen in a way that average earnings are similar across treatments.

For the punishment treatments of experiment two, we deduct 2 Euro from the payoff of a participant whenever her color is selected by the evaluator. For the reward treatments, we add 2 Euro to the payoff of a participant whenever her color is selected by the evaluator.

Participants in the role of an evaluator receive a flat payment of 16 Euros. In the treatments with correlation incentives, the evaluators receive 2 Euros whenever both evaluators select the same group member in the same round.

5 Experimental results

5.1 Response to social influence

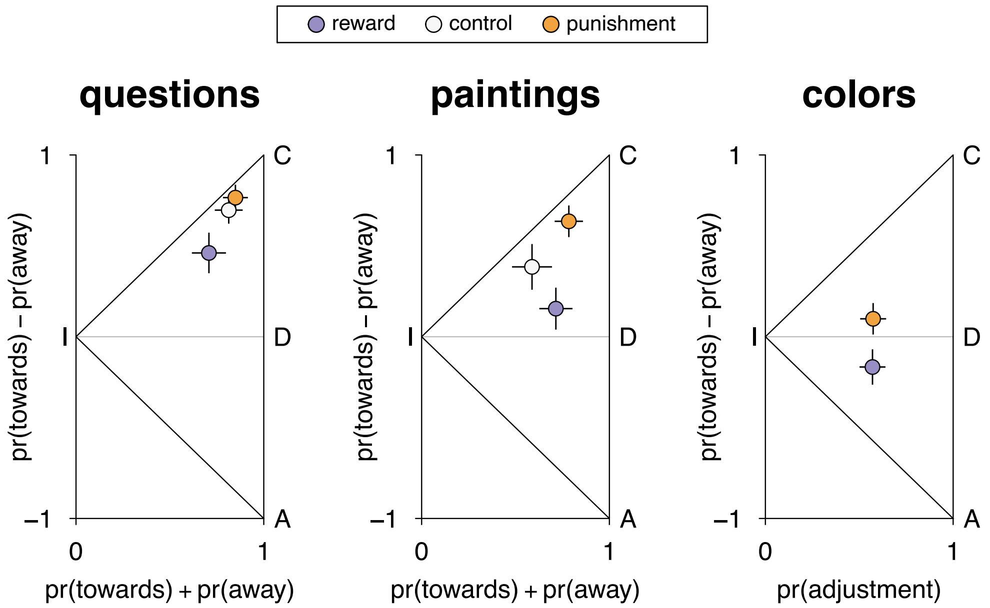
We illustrate the effect of punishment and reward on the response to social influence by pooling the data of the treatments with and without coordination incentives. Figure B.1 in Appendix B shows that both treatment variants generate similar responses to social influence.

Average response to social influence

The colored dots in Figure 3 depict the average response to social influence in the punishment, reward and control treatments. Whiskers indicate the 95 percent confidence interval of the average response along the two model dimensions based on the t-distribution and block-bootstrapped standard errors.

The right panel depicts the average response in the second half of experiment two (periods 5-8) using the minimal distance variable and the rgb color metric. We depict the results for the second half of experiment two, because after period 4 participants gained sufficient experience with the color creation interface and created a desirable variety of colors which allows them to react to others' choices. Figure B.2 in Appendix B shows that the treatment effects are robust if we use data from all periods or other distance functions to calculate the coordinates of the average response to social influence in experiment two.

Figure 3: Average response to social influence across treatments



Notes: Dots show average behavior, whiskers 95 percent confidence interval based on block-bootstrapped standard errors. Blocks are subjects in experiment one and matching groups in experiment two. Results of the colors panel for minimal distance in rgb colorspace in the second half of the experiment (periods 5-8).

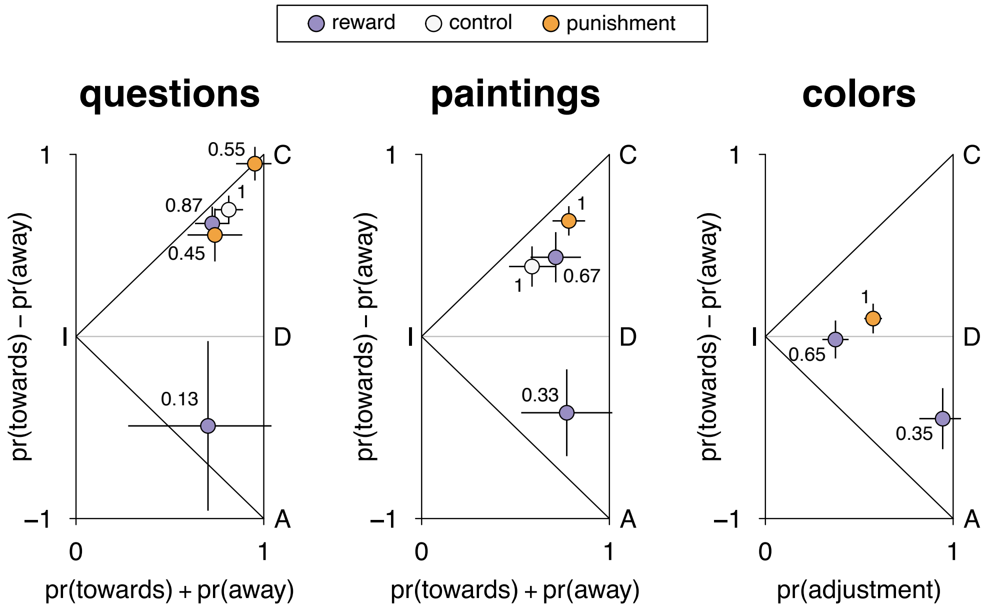
Figure 3 contains three results. First, the dots in the left and central panels are generally closer to the conformity vertex of the triangle. This means we observe a higher level of conformity in the questions and paintings tasks compared to the color task. Second, the differences across the experimental treatments are in line with the comparative statics of our theoretic model. Participants exhibit less conformity in the reward treatments compared to the control and punishment treatments in the questions and paintings tasks. In the colors task, we observe that average behavior shifts from the domain of conformity with punishment to the domain of anticonformity with reward. Third, the position of average behavior on the independence-dependence dimension is not consistently affected by the treatments.

Heterogeneity in the response to social influence

Figure 4 depicts individual differences in the response to social influence. The colored dots indicate the response to social influence of a behavioral type. The position and the prevalence of each behavioral type in the data of a treatment is estimated on the basis of a mixture model. The number of types of the mixture model is identified for the data of each treatment based on a five-fold cross-validation procedure. The number next to each dot shows the estimated prevalence of the type in the sample. Whiskers indicate the block-bootstrapped 95 percent confidence interval of the response to social influence of a type for the two model dimensions.

Figure 4 reveals individual differences in the response to social influence. For the

Figure 4: Individual differences in the response to social influence



Notes: Dots indicate average behavior of types, numbers next to the dots the estimated prevalence of the type in the sample. Whiskers indicate the 95 percent confidence interval for each type based on block-bootstrapped standard errors. Number of types is identified for each treatment by k-fold cross validation (k = 5). See Appendix B.2 for details.

reward treatments of all three tasks, the cross-validation indicates two types which substantially differ in their position of the conformity-anticonformity dimension. The more frequent type with estimated frequencies ranging from 0.65 to 0.87 is conformist in the questions and paintings task and rather independent in the colors task. The minority type with estimated frequencies ranging from 0.13 to 0.35 is clearly anticonformist with a similar degree of anticonformity across all three tasks.

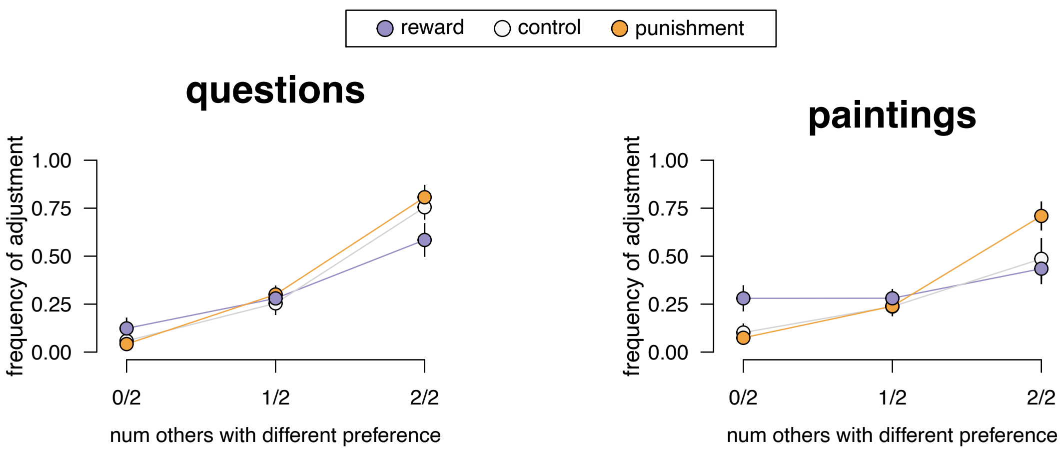
Apart from the heterogeneity in the reward treatments, the cross-validation also indicates heterogeneous responses in the punishment treatments of the questions task. The heterogeneity is characterized by a close to uniform mixture of two types. One type is moderately conformist while the other is extremely conformist and adjusts to the majority choice almost every time when this is possible. For all other treatments, the cross-validation procedure indicates no heterogeneity in the response to social influence. The estimated position of the single behavioral type therefore corresponds to the average social response depicted in Figure 3.

For the mixture models with two types, we investigate whether the prior probability to belong to a certain type can be explained by individual traits of the participants. We fit a latent class regression model (Bandeen-Roche et al., 1997) which models the prior probability to belong to one of the two types as a function of the individual sum score of participants' answers in a post-experimental conformity questionnaire (Mehrabian and Stefl, 1995, see Appendix B.2 for details). The estimated coefficients of the conformity sum score have the expected sign in each of the four latent class regression models. The prior probability to belong to the more conformist type increases with a participants' conformity score. The relationship is significant in the reward treatments of the paintings task. In these treatments, scoring one standard deviation above rather than below the average conformity score of the sample increases the prior probability to belong to the conformist type from 49 to 85 percent (Table B.2 in Appendix B.2).

Social choices in experiment one

Figure 5 shows the relative frequency of an adjustment of the social choice across three different scenarios of experiment one. The three different scenarios differ in the number of other group members that did not choose in line with the prediction for the social choice. The colored dots indicate the relative frequency of adjustment of the social choice if the choices of zero, one or two other group members indicate the other preference. If none out of two other group members indicate the other preference, the participant can adjust the social choice away from the majority. The target participant can adjust the social choice towards the majority if two out of two other group members indicate the other preference. The scenario in between marks the absence of social influence, as the choices of the other two group members differ. The steeper the frequency of adjustment increases in the number of other group members with the other preference, the more conformist the adjustment strategy is.

Figure 5: Adjustment of social choices in experiment one



Notes: Dots show the relative frequency of an adjustment of the social choice conditional on the number of other group members with a different preference in experiment one. Whiskers indicate 95% CIs of the estimates, based on block-bootstrapped standard errors (1000 samples, matching subject ID).

Figure 5 unveils that the treatment differences in the average response to social influence in the questions task depicted in Figure 3 arise from a lower frequency of adjustment towards the others' choices in the reward treatment. This means that the social choices in the reward treatments are less conformist if the choices of the other two group members differ from the own preference. In the paintings task, we observe two notable differences in the frequency of adjustment across treatments.

Participants more often adjust the social choice towards others' choices in the punishment treatments, and away from others' choices in the reward treatments.

Figure 5 also shows that the frequency of adjustment of the social choice in the absence of social influence does not differ across treatments. This indicates a comparable quality of the predictions of the social choices across treatments.

The theoretical model suggests that the social choices in the reward treatments are adjusted conditional on the difficulty of the choice. If the choice is difficult, participants should more (less) frequently adjust away from (towards) the majority. In the theoretical model, choice difficulty is the share p of participants which prefers the less popular alternative. For the experimental data, we estimate the difficulty of a choice by the share of participants which chooses the less popular alternative in the absence of social information. We correlate the estimated choice difficulty to the relative frequencies of adjustments in the two situations with social influence. The Pearson correlation coefficient of the relative frequency of adjustments towards the majority and choice difficulty is negative (-0.06, bootstrapped 95% CI: -0.15, 0.02). The correlation coefficient of the relative frequency of adjustments away from the majority and choice difficulty is positive (0.11, bootstrapped 95% CI: 0.05, 0.17).

Social choices in experiment two

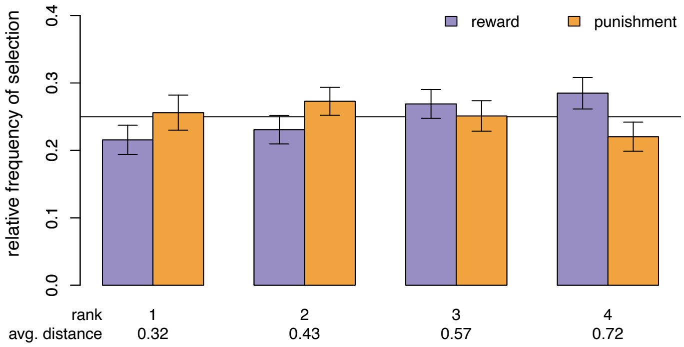
If a participant adjusts the social choice in the colors task, she can choose up to three different degrees of similarity to the colors of the other group members. The number of different degrees of similarity depends on the four colors she can chose from in the current round. The vertical position of the average response to social influence depicted in Figure 3 indicates the average direction of the adjustments. However, it does not reveal the degree of similarity between the adjusted color and the colors of the other group members after an adjustment.

To investigate the degree of similarity, we calculate the average distance of the adjusted colors to the colors of the other three group members in the reward and punishment treatments. We find a higher degree of similarity of the adjusted colors to the colors of the other three group members in the punishment treatments. Table B.1 in Appendix B shows that the average distance of the adjusted colors in the punishment treatments is consistently smaller for all six distance measures. Figure B.4 in Appendix B illustrates that the average distance decreases over time as participants gain experience over the course of the experiment.

The difference in the degree of similarity of the adjusted colors between the punishment and reward treatments is driven by different adjustment strategies across the treatments. To identify the different adjustment strategies, we calculate the rank of each color that could be used for the adjustment to the colors of the other group members. We assign the highest rank to the color with the largest minimum Euclidean distance to the colors of the other three group members.

Figure 6 compares the frequency of adjustment ranks across the reward and punishment treatments. In the reward treatments, participants most frequently adjust to the color with the largest distance rank. In the punishment treatments, participants most frequently adjust to the colors with ranks that indicate a smaller distance to

Figure 6: Adjustment to distance ranks in experiment two



Notes: Relative frequency of distance ranks. Higher ranks indicate a larger minimum Euclidean distance of the selected color. Whiskers indicate 95% CIs, based on block-bootstrapped standard error of the mean (1000 samples, matching group ID).

the colors of the other participants.

Determinants of the adjustment of social choices

Next, we identify factors which influence the decision to adjust the social choice. We regress a dummy indicating an adjustment on characteristics of the predicted nonsocial choice. Table 3 depicts the coefficients of logit models fitted to the data of the experimental treatments with block-bootstrapped standard errors in parentheses.

The upper part of the table contains the coefficients and standard errors of the covariates we use in the logit models for experiment one. The coefficients show that the decision to adjust the social choice is influenced by two aspects.

The first aspect which influences the decision to adjust in the questions and paintings task is the predicted preference strength of the nonsocial choice. The variable preference strength is the predicted strength of the preference (or judgment) for the predicted nonsocial choice. The coefficients of the predicted preference strength are consistently negative across all treatments in both tasks. This indicates that participants rarely adjust the social choice if the preference for the predicted choice is strong.

The second aspect which influences the decision to adjust the social choice in the questions and paintings task concerns the different incentives that arise from social evaluations across treatments. The variables majority choice and unique choice are dummies which indicate if the predicted nonsocial choice is chosen by both other group members or by no other group member. The negative coefficients of the dummy majority choice reflect the conformist response to social influence in both tasks across all experimental treatments. The positive coefficients of the dummy

Table 3: Logit models for the adjustment of choices

	questions			paintings			colors	
	reward	control	punish	reward	control	punishment	reward	punish
preference strength	-1.03	-1.35	-1.59	-1.28	-2.69	-1.26	-	-
	(0.40)	(0.40)	(0.42)	(0.27)	(0.55)	(0.36)	-	-
majority choice	-1.61	-2.41	-2.64	-0.38	-1.22	-2.04	-	-
	(0.27)	(0.34)	(0.33)	(0.19)	(0.30)	(0.27)	-	-
unique choice	0.69	1.54	1.99	0.17	0.87	1.39	-	-
	(0.22)	(0.23)	(0.24)	(0.19)	(0.28)	(0.21)	-	-
intercept	-	-	-	-	-	-	0.71	0.41
	-	-	-	-	-	-	(0.25)	(0.23)
min distance	-	-	-	-	-	-	-0.65	0.02
	-	-	-	-	-	-	(0.25)	(0.23)
beauty rating	-	-	-	-	-	-	-0.23	-0.02
	-	-	-	-	-	-	(0.24)	(0.29)
interest rating	-	-	-	-	-	-	0.02	0.06
	-	-	-	-	-	-	(0.27)	(0.30)
Obs	571	356	607	569	360	590	1376	1408
N	105	60	105	105	60	105	172	176

Notes: Table shows logit coefficients and block-bootstrapped standard errors in parentheses. Dependent variable is a dummy which indicates if the social choice is adjusted. The symbol '-' indicates that the covariate is not included in the model depicted in this column. *Obs* and *N* indicate is the number of observations and participants in the sample.

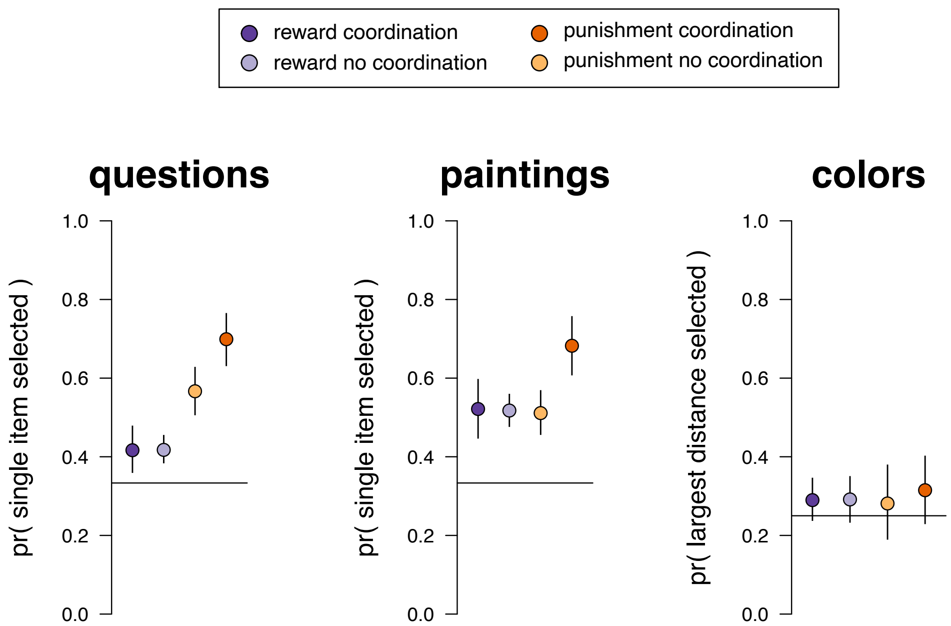
unique choice indicate that participants frequently deviate from the predicted alternative if the alternative is chosen by no other group member. The size of the coefficients systematically differs across treatments in a way predicted by theory. The probability to adopt the majority choice is smallest in the reward treatments and highest in the punishment treatments.

The lower part of Table 3 contains the coefficients and standard errors of the covariates we use in the logit models for experiment two. For the data of the colors task, the logit coefficients suggest that the decision to adjust the choice under social influence is affected by strategic considerations, but only in the reward treatment. Participants more frequently adjust in the reward treatment if the minimum of the Euclidean distances of the nonsocial choice to the other colors is small, which means that the own color is similar to a color of another participant. Post-experimental ratings of the participants of how beautiful and interesting the created colors are do not influence the decision to adjust the social choice.

5.2 Evaluation

As the evaluation of choices differs between the treatments with and without coordination incentives, we report the results for each treatment separately.

Figure 7: Expected benefit and cost of standing out



Notes: Dots indicate the probability to select the single answer or question in the left and central panel. Dots indicate the probability of selecting the color with the largest minimal distance in right panel. Whiskers indicate block-bootstrapped 95 percent confidence intervals. Horizontal lines indicate the expected probability if evaluators would randomly allocate punishment and reward.

Incentives for standing out

Figure 7 compares the incentives to stand out in the three experimental tasks across the four treatments with social evaluation. The left panel of Figure 7 shows the relative frequency of selecting the answer or painting which stands out as it is chosen by one group member. The right panel shows the relative frequency that the evaluator selects the color with the largest minimal distance to the colors of the other group members.

The frequency of selecting the single answer or painting is clearly larger than one third in the questions paintings tasks in all treatments with social evaluation. This implies that the participant who stands out is most frequently rewarded or punished. In addition, the left and central panel of Figure 7 highlight a notable asymmetry in the effect of coordination incentives for the evaluators. The frequency of selecting the single question or answer is higher with coordination incentives but only in combination with punishment. In combination with reward, coordination incentives have no effect on the frequency of selecting the single question or answer.

The right panel of Figure 7 reveals that the relative frequency that the evaluator selects the color with the largest minimal distance is larger than one fourth. One fourth is the relative frequency we would expect if evaluators would randomly pick one of the four colors. Compared to the results for the questions and paintings task, the incentive for standing out in the color task are smaller. The right panel also shows that the incentives for standing out are of the same magnitude in the presence and absence of coordination incentives.

Taste-based vs. salience-based evaluation

Table 4 depicts the results of conditional logit models (McFadden, 1974) to disentangle taste-based and salience-based allocation of punishment and reward. In the conditional logit model, the choice over several alternatives is modeled as a function of the characteristics of the alternatives. To investigate whether reward and punishment are allocated based on taste or based on salience we regress a dummy which indicates the choice selected by the evaluator 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 models fitted on the data of experiment one contain a dummy which indicates the single answer or painting, and a dummy indicating the preference of the evaluator in the same choice. For the model we fit on data of the colors task, we elicit the color which is generally considered to be salient by performing a Krupka-Weber coordination task (Krupka and Weber, 2013) over the four colors after the experiment. The model includes a dummy for the salient color and the evaluator's rating how beautiful and interesting each color is using two continuous scales ranging from zero (not beautiful / not interesting at all) to one (very beautiful / very interesting).

Table 4 shows that taste plays an important role for the allocation of reward and punishment. The coefficients of the models which can be related to taste (preferred choice, beauty and interest rating) are generally positive for the reward treatments

Table 4: Taste-based vs. salience-based evaluation

	questions				paintings				colors			
	reward		punishment		reward		punishment		reward		punishment	
	nc	c	nc	c	nc	c	nc	c	nc	c	nc	c
salient choice	-0.50	-0.30	0.26	0.93	0.10	0.07	0.14	0.76	-	-	-	-
	(0.09)	(0.10)	(0.09)	(0.08)	(0.09)	(0.08)	(0.09)	(0.08)	-	-	-	-
preferred choice	1.53	1.40	-1.07	-0.44	1.21	1.12	-1.26	-0.57	-	-	-	-
	(0.13)	(0.13)	(0.12)	(0.11)	(0.12)	(0.11)	(0.13)	(0.11)	-	-	-	-
salient color	-	-	-	-	-	-	-	-	1.11	1.25	0.25	1.17
	-	-	-	-	-	-	-	-	(0.19)	(0.15)	(0.23)	(0.17)
beauty rating	-	-	-	-	-	-	-	-	1.09	1.91	-2.15	0.35
	-	-	-	-	-	-	-	-	(0.56)	(0.42)	(0.58)	(0.40)
interest rating	-	-	-	-	-	-	-	-	0.56	1.07	-1.22	0.25
	-	-	-	-	-	-	-	-	(0.56)	(0.48)	(0.91)	(0.48)
Obs	648	612	648	612	612	648	612	648	736	1280	768	1280
N	54	51	54	51	51	54	51	54	23	40	24	40
LL	-302	-303	-366	-357	-342	-349	-367	-342	-213	-308	-234	-383

Notes: Conditional logit coefficients and block bootstrapped standard errors in parentheses. Labels 'nc' and 'c' indicate the treatments without and with coordination incentives. Variables *salient choice* and *preferred choice* are dummies which indicate the single and own choice. *min distance* is the minimum of the Euclidean distances to the other three colors. beauty and interest ratings of the colors were elicited after the experiment on a scale from zero (low rating) to one (high rating). Labels *Obs*, *N* and *LL* indicate is the number of observations, participants, and the log likelihood of the model.

and negative for the punishment treatments. This indicates that participants frequently allocate reward to other participants with the same preferences or opinions and punishment to those with different preferences or opinions. An exception to this pattern is the punishment treatment in the color task with coordination incentives. In this treatment, beautiful and interesting colors do not escape to be selected for punishment, suggesting a minor role of tase-based allocation.

The evidence for salience-based allocation of reward and punishment is mixed. If punishment and reward were allocated based on salience, the coefficients of the variables reflecting salience (salient choice, salient color) should be positive. This would indicate that more salient choices are more frequently selected. The coefficients of the salience variables depicted in Table 4 indicate that salience plays a minor role in the absence of coordination incentives (columns labeled 'nc'). In the treatments without coordination incentives, the coefficients are often small and insignificant. In the reward treatments of the questions task, the coefficients are negative which means that the single choice is less likely to be selected for reward. One explanation for the negative coefficients could be that evaluators try to reward correct answers and believe that the majority is probably correct. This is clearly not in line with the role salience plays in our model. However, the coefficients for salience show the hypothesized sign in the treatments with coordination incentives (columns labeled 'c'), especially when punishment is the consequence of selection.

6 Discussion

The idea that social evaluation can induce strategic conformity has a long history in the social sciences. Social psychologists argue that people conform to others' preferences and judgments in order to attract reward and avoid punishment (Festinger, 1953; Kelman, 1961; Allen, 1965). In the present study, we theoretically and experimentally investigate the effect of social evaluation on conformity and anticonformity. Based on a theoretical model, we illustrate that the social evaluation of choices creates incentives for strategic conformity if the objective of the evaluation is to select one individual of a group for punishment. Yet, we also illustrate that the social evaluation of choices creates incentives for strategic anticonformity if the objective of the evaluation is to select one individual of a group for reward.

In the model, conformity is strategic as it decreases the probability of being punished. Anticonformity is also strategic as it increases the probability of being rewarded. In two laboratory experiments, we find strategic (anti)conformity in publicly disclosed preferences, judgments and decisions in a creativity-related task. Social evaluation of choices induces strategic conformity in case of potential punishment, and strategic anticonformity in case of potential reward.

The average response to social influence varies considerably across the three different experimental tasks. We observe the highest degree of conformity in the task in which participants select answers to questions that have an objectively correct answer. We find a lower degree of conformity in the task in which participants select art paintings, where preferences are entirely subjective. Since the questions are rather difficult, the finding of the highest degree of conformity in the questions task is in line with studies which observe that conformity increases with task difficulty (Baron et al., 1996). Our result contrasts however with previous experimental findings showing that conformity is attenuated in tasks with an objectively correct answer (Kameda et al., 2003). We find the lowest degree of conformity in the creativity-related task, in which participants choose from several alternatives that vary along multiple dimensions.

We also find substantial individual variation in the response to social influence. Variation in the response to social influence is a pattern which pervades the literature on individual differences in social learning in humans and other species (see Mesoudi et al., 2016, for a review). Across all three experimental tasks, a minority of participants in the reward treatments displays a strong anticonformist response to social influence. In contrast, the majority of participants does not follow the monetary incentives for anticonformity.

One potential explanation for this finding is that some participants derive substantial intrinsic utility from conforming to others' choices. The control treatments of experiment one lend some support to this interpretation as the average response to social influence is conformity in the absence of social evaluation. The monetary incentives for anticonformity which exist with social evaluation might thus not be strong enough for some participants to adjust their behavior. We also find some support for the interpretation that these participants differ in individual traits related to conformity.

Yet, another potential explanation for the fact that not all participants exploit the monetary incentive for anticonformity might be that some participants do not realize that incentives for anticonformity exist. In experiment one, this requires that the participants correctly anticipate the allocation of reward. In experiment two, participants need to draw the correct conclusion from the observation of the allocation of reward in previous periods. Both is not trivial.

The finding that only few participants display anticonformity in our experiments raises the question whether anticonformity by a minority is an important phenomenon. Yet, even a single anticonformist response can break unanimity which is a particularly strong determinant of conformity in subsequent choices (Asch, 1955). A single anticonformist response can also break an established information cascade which might explain why information cascades are not as frequent as predicted by theory in laboratory experiments (Huck and Oechssler, 2000; Nöth and Weber, 2002). It has also been shown that anticonformity by a minority can play an important role for the formation of the majority opinion and in the polarization of opinions (Juul and Porter, 2019; Siedlecki et al., 2016). Finally, findings from statistical physics explain why anticonformity is usually rare. An anticonformist majority cannot exist in the thermodynamic limit of a system of interacting agents in which the agents react to each other with some delay (Touboul, 2019). In contrast, the delay in the responses to social influence induce synchronized oscillations in anticonformist behavior such that anticonformists follow the majority behavior at every point in time.

Our experiments also shed light on how individuals evaluate others' behavior. In the experiments, we find that participants allocate reward and punishment mainly based on taste. Across all experimental tasks, participants tend to allocate reward to those who publicly display similar preferences or judgments and punishment to those who publicly display different preferences or judgments. We observe allocation of reward and punishment based on salience when incentives for the coordination of evaluation decisions exist. Yet, the taste-based allocation of punishment and reward prevails even in the presence of such incentives as the evaluators seem to not reach a consensus on what is salient.

The notion that people prefer to allocate benefits to individuals with similar preferences and costs to individuals with different preferences is supported by the psychological literature on ingroup favoritism (see Hewstone et al., 2002, for a review). The same idea is also prominently discussed in the economic literature on taste-based discrimination (Becker, 1971; Riach and Rich, 2002; Bertrand and Duflo, 2017).

Our theoretical model highlights three important conditions which must be fulfilled in order to induce strategic anticonformity on the basis of social evaluation. First, the consequences of social evaluation affect a subset of the evaluated individuals. Second, the preferences and judgments of the individuals which evaluate behavior are unknown. Third, the evaluated individuals believe that preferences or judgments vary to some extent between evaluators.

These three conditions occur naturally in everyday life. For example, the consequences of social evaluation do often affect only a subset of individuals because the implementation of consequences is costly for the evaluating individual. In other cases, the number of individuals affected by the consequences of social evaluation is regulated by institutions.

At the same time, the preferences and judgments of the individuals which evaluate our behavior are often unknown. In modern societies people frequently interact with strangers. The plurality of preferences and opinions in society can create uncertainty about the attitudes of people we interact with on a daily basis.

Finally, the condition that the evaluated individuals must believe that preferences or judgments vary in the population is interesting for the evolution of behavior over time. The critical prior probability identified in the theoretical analysis for which punishment induces strategic conformity and reward induces strategic anticonformity can be interpreted as the tipping points in the dynamic model of behavior. If individuals use the frequency of observed behavior in the population in a naive way as the belief for the distribution of preferences, the entire behavioral system moves to unanimity when crossing these thresholds. Independent of the consequences of social evaluation, the system cannot escape the state of perfect unanimity which highlights the role of diversity for sustaining the plurality of opinions.

From the perspective of society, the benefit of conformity and anticonformity can vary considerably across situations. While conformity can be the reason why people adhere to social norms it can also be the reason why people make irrational financial decisions, shy away from innovative practices, are susceptible to group think, and communicate in echo chambers or filter bubbles. The same applies to anticonformity, which can help to start innovation, break undesirable routines and erode archaic social conventions but at the same time reduces coordination and predictability of behavior, which can be detrimental for society. In this context, the potential of social evaluation to shift the average response to social influence looms large.

A Theoretical Appendix

Preliminaries

Without loss of generality we investigate the incentive for A_1. We fix the probability q_A to follow the own taste for players A_2 to A_{N-1} and the strategy of player B. We set $p_A := pq_A + (1-p)(1-q_A)$. Because $p > \frac{1}{2}$ and $q_A \geq \frac{1}{2}$, we get $p_A > \frac{1}{2}$. It is the probability that A chooses according to the commonly preferred option. Let q_1 be player A_1's probability to follow the own taste, and $p_1 = pq_1 + (1-p)(1-q_1)$ be player A_1's probability to choose according to the commonly preferred option. Further, we define $B(n,k,p) := \binom{n}{k} p^k \cdot (1-p)^{n-k}$ for the binomial distribution.

Evaluation based on salience

Proposition 1 (Salience under punishment)**.** *The A players follow their taste. If there is a strict majority, i.e. a majority of size $> \frac{(N-1)}{2}$, then B follows this majority, and decides according to the own taste otherwise.*

Proposition 2 (Salience under reward)**.** *The A players randomize with a probability of $\frac{1}{2}$. If there is a strict minority, i.e. a minority of size $< \frac{(N-1)}{2}$, then B follows this minority (which can be empty), and decides according to the own taste otherwise.*

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 signal. If there is an incentive for anticonformity, then A players should discoordinate, which they best achieve by randomizing.

We give the formal proof for the case of reward. The case of punishment can be shown analogously.

Proof salience-based reward. For notational convenience, we set $m = \frac{N-1}{2}$. Without loss of generality we investigate the incentive for A_1. We fix the probability q_A to follow the own taste for players A_2 to A_{N-1} and the strategy of player B.

Consider an odd number of players. The expected winning probability W for player A_1 equals (note than in the majority the probability to win equals 0, and $p, q_A \geq \frac{1}{2}$)

$$
\begin{aligned}
W = p_1 &\left(\sum_{k=0}^{m-2} \frac{B(N-2,k,q_A)}{k+2} + (1-p)\frac{B(N-2,m-1,q_A)}{m} \right) + \\
&+ (1-p_1) \left(\sum_{k=0}^{m-2} \frac{B(N-2,k,1-q_A)}{k+2} + p\frac{B(N-2,m,1-q_A)}{m} \right) \\
&= \sum_{k=0}^{m-2} \frac{B(N-2,k,q_A) - B(N-2,k,1-q_A)}{k+2} + \\
&(1-p)\frac{B(N-2,m-1,q_A)}{m} - p\frac{B(N-2,m-1,1-q_A)}{m} < 0
\end{aligned}
$$

Thus, it is optimal for A to to choose against the own taste – conditional on $q_A \geq \frac{1}{2}$, which mans that $q_A = \frac{1}{2}$.

Now, consider an even number of players. With an even number of players, one of the N players is randomly selected if each alternative is chosen by $N/2$ players. For this case, we set $m = \frac{N}{2} - 1$. The expected winning probability W for player A_1 equals

$$
\begin{aligned}
W = p_1 &\left(\sum_{k=0}^{m-2} \frac{B(N-2,k,q_A)}{k+2} + \sum_{k=m-1}^{m} \frac{B(N-2,m-1,q_A)}{N} \right) + \\
&+ (1-p_1) \left(\sum_{k=0}^{m-2} \frac{B(N-2,k,1-q_A)}{k+2} + \sum_{k=m-1}^{m} \frac{B(N-2,m,1-q_A)}{N} \right) \\
&= \sum_{k=0}^{m-2} \frac{B(N-2,k,q_A) - B(N-2,k,1-q_A)}{k+2} + \\
&\sum_{k=m-1}^{m} \frac{B(N-2,m-1,q_A) - B(N-2,m-1,1-q_A)}{N} < 0
\end{aligned}
$$

Again, it is optimal for A to to choose against the own taste which mans that $q_A = \frac{1}{2}$. □

Evaluation based on taste

Proposition 3 (Taste based punishment)**.** *Independent of the strategy of player B, the A players always follow their taste. B is always conformist; if there is a strict majority, i.e. a majority of size >* $\frac{(N-1)}{2}$*, then B follows this majority, and decides according to the own taste otherwise.*

Proof of taste based punishment proposition. We first show, that B is never anti-conformist. Using this, we will show that $q_A = 1$, and then that B conforms. As in the proof of the reward proposition below, there is also a direct proof that shows $q_A = 1$. We prefer this proof because it is more intuitive.

We start showing that B is never anticonformist. First, the probability p_A that A's choice correspond to the generally preferred alternative is at least $\frac{1}{2}$ because $q_A \geq \frac{1}{2}$ and $p > \frac{1}{2}$. This means that anticonformity is not advantageous advantageous with respect to the evidence that the A players provide with their decision. Second, anticonformity goes against the evidence from the own taste. Third, joining the (weak) minority also leads to a (weakly) higher probability to be punished if the chosen options turns out to be the generally less preferred because in this case the punishment probability is shared among fewer people. Thus, B is never anticonform. This means that if B does not follow the own taste, B follows the majority.

Now turn to A. Let A_1 be the focal player. As player B, player A_1 has an incentive to be correct or to be in the majority. Following the own taste is the only information about correctness. Since $q_A \geq \frac{1}{2}$ in expected terms, more that half of the other A players make a choice according to the own taste. In addition, signaling the own taste to B has the advantage to potentially increasing the majority, because B is never anticonformist. Thus, A should follow the own taste.

Now, because $q_A = 1$ the evidence of the majority – also together with a contradicting evidence of player B is evidence that this choice has a probability higher than $\frac{1}{2}$. In addition, joining the majority reduces the risk of being punished if this option is the option preferred by the evaluator. Thus, B's optimal choice is to conformist. $\square$

Proposition 4 (Taste based reward)**.** *If $N = 3$, the two A players always follow their taste, independent of the strategy of player B. For $k = 0$, Player B's choice is conformist iff p is above some threshold $C(N,k) < 1$. Player B's choice is anticonformist iff $p > \frac{1}{2} + \frac{\sqrt{3}}{6}$. For $k = 2$, Player B's choice is anticonformist iff $p < \frac{1}{6}\left(3 + \sqrt{6\sqrt{3} - 9}\right)$. For $k = 1$, Player B's choice is independent.*

We start with three lemmas which apply for all values of q_A and characterize the best response of B.

Lemma 1. *For $k = 1$, B follows the own taste.*

Lemma 2. *For $k = 0$, B follows the own taste if $p \leq 2/3$.*

Lemma 3. *For $k = 2$, B is anticonformist if $p < 2/3$.*

Then we study the effect of B's behavior on the incentives of the focal player A_1 to follow the own taste. This is done based on the derivative of the probability to win the reward W with respect to the probability to choose the own taste p_1. Using Lemmas 1-3, we show that $\partial W/\partial p_q$ is positive for $p < 2/3$ and all possible values of q_A.

For the case $p \geq 2/3$ we use two additional lemmas which specify the behavior of B for $k \neq 1$ and $p \geq 2/3$ that minimizes the incentive of A_1 to follow the own taste:

Lemma 4. *For $k = 0$ and $p \geq 2/3$, conformity minimizes $\partial W/\partial p_1$.*

Lemma 5. *For $k = 2$ and $p \geq 2/3$, independence minimizes $\partial W / \partial p_1$.*

Using Lemmas 1, 4 and 5, we show that $\partial W / \partial p_q$ is positive for $p \geq 2/3$ and all possible values of q_A. This proves the first statement of Proposition 4. To complete the proof of Proposition 4, we specify the best response of B in equilibrium given that $q_1 = q_A = 1 \rightarrow p_1 = p_A = p$. For $k = 0$, B is conformist if $p > \frac{1}{2} + \frac{1}{2\sqrt{3}}$. For $k = 2$, B is anticonformist if $p < \frac{1}{6}\left(3 + \sqrt{6\sqrt{3} - 9}\right)$. This proofs Proposition 4.

Behavior of B

Let $\beta : K \rightarrow \{0, 1\}$ denote a function that indicates if player B chooses the own taste given that $k \in K = \{0, 1, 2\}$ A players choose this taste. Let p^H denote the probability that the taste of player B is generally preferred. After observing the choices of the A players, B updates p^H according to:

$$p^H = \frac{pB(k, 2, p_A)}{pB(k, 2, p_A) + (1 - p)B(k, 2, 1 - p_A)}$$

Lemma 1. *For $k = 1$, $\beta(1) = 1$.*

Proof. For $k = 1$, $p^H = p$ and B follows the own taste if:

$$\frac{p^H p + (1 - p^H)(1 - p)}{2} \geq \frac{p^H(1 - p) + (1 - p^H)p}{2},$$

which yields

$$p^2 + (1 - p)^2 \geq 2p(1 - p),$$

and is always true if $p > 0.5$. This shows that B follows the own taste which means $\beta(1) = 1$. □

Lemma 2. *For $k = 0$, $\beta(0) = 1$ if $p \leq 2/3$.*

Proof. For $k = 0$, B follows the own taste if:

$$p^H p + (1 - p^H)(1 - p) \geq \frac{1}{3},$$

which yields

$$p - p^H(2p - 1) \leq \frac{2}{3}, \tag{9}$$

For $p^H \in (0, 1)$, this is always fulfilled if $p \leq 2/3 \rightarrow \beta(0) = 1$ if $p \leq 2/3$. □

Lemma 3. *For $k = 2$, $\beta(2) = 0$ if $p < 2/3$.*

Proof. For $k = 2$, B follows the own taste if:

$$\frac{1}{3} \geq p^H(1 - p) + (1 - p^H)p,$$

which yields

$$\frac{2}{3} \leq p - p^H(2p - 1), \tag{10}$$

For $p^H \in [0, 1]$, this is never fulfilled if $p < 2/3 \rightarrow \beta(0) = 0$ if $p < 2/3$. □

Effect of B's behavior on A_1

Now, we investigate the effect of B's behavior on the incentive of the focal player A_1 to follow the own taste. Using Lemma 1, the probability that the focal player A_1 receives the reward is:

$$\begin{aligned}
W = {} & p_1 p_A p \left(\frac{\beta(2)}{3} + \frac{p(1-\beta(2))}{2} \right) \\
& + p_1 p_A (1-p) \left(\frac{1-\beta(0)}{3} + \frac{p\beta(0)}{2} \right) \\
& + p_1 (1-p_A) p \frac{p}{2} \\
& + p_1 (1-p_A)(1-p) p \\
& + (1-p_1) p_A p (1-p) \\
& + (1-p_1) p_A (1-p) \frac{1-p}{2} \\
& + (1-p_1)(1-p_A) p \left(\frac{(1-p)\beta(0)}{2} + \frac{1-\beta(0)}{3} \right) \\
& + (1-p_1)(1-p_A)(1-p) \left(\frac{\beta(2)}{3} + \frac{(1-p)(1-\beta(2))}{2} \right)
\end{aligned}$$

The incentive of the focal player A_1 to follow the own taste is reflected by the partial derivative of the winning probability with respect to p_1. The partial derivative $\partial W / \partial p_1$ is:

$$\begin{aligned}
\frac{\partial W}{\partial p_1} = {} & p_A p \left(\frac{\beta(2)}{3} + \frac{p(1-\beta(2))}{2} \right) \\
& + p_A (1-p) \left(\frac{1-\beta(0)}{3} + \frac{p\beta(0)}{2} \right) \\
& + (1-p_A) p \frac{p}{2} \\
& + (1-p_A)(1-p) p \\
& - p_A p (1-p) \\
& - p_A (1-p) \frac{1-p}{2} \\
& - (1-p_A) p \left(\frac{(1-p)\beta(0)}{2} + \frac{1-\beta(0)}{3} \right) \\
& - (1-p_A)(1-p) \left(\frac{\beta(2)}{3} + \frac{(1-p)(1-\beta(2))}{2} \right)
\end{aligned} \tag{11}$$

$\partial W / \partial p_1 > 0$ **for** $p < 2/3$

Proof. For $p < 2/3$, Lemmas 1-3 show $\beta(2) = 0$ and $\beta(0) = 1$. Given this behavior of B, the probability that the focal player A_1 receives the reward specified in

Equation 11 becomes:

$$\begin{aligned}\frac{\partial W}{\partial p_1} &= p_A p \frac{p}{2} \\ &+ p_A(1-p)\frac{p}{2} \\ &+ (1-p_A)p\frac{p}{2} \\ &+ (1-p_A)(1-p)p \\ &- p_A p(1-p) \\ &- p_A(1-p)\frac{1-p}{2} \\ &- (1-p_A)p\frac{1-p}{2} \\ &- (1-p_A)(1-p)\frac{1-p}{2}\end{aligned}$$

which boils down to

$$\frac{\partial W}{\partial p_1} = \frac{1}{2}\left(p(3-p) - 1 - p_A(1-p)2p\right)$$

which is greater or equal to zero for $0.5 \leq p$ and $p_A \leq p$.

□

$\partial W/\partial p_1 > 0$ **for** $p \geq 2/3$

To prove the case $p \geq 2/3$, we use the additional lemmas which characterize the behavior of B that minimizes $\partial W/\partial p_1$.

Lemma 5. *For $k = 0$ and $p \geq 2/3$, conformity minimizes $\partial W/\partial p_1$.*

Proof. Conformity decreases $\partial W/\partial p_1$ if:

$$p_A\frac{(1-p)}{3} \; - \; (1-p_A)\frac{p}{3} < p_A(1-p)\frac{p}{2} \; - \; (1-p_A)p\frac{(1-p)}{2}$$

which gives

$$p_A(1-p)\left(\frac{1}{3} - \frac{p}{2}\right) \; + \; (1-p_A)p\left(\frac{1-p}{2} - \frac{1}{3}\right) < 0$$

and is fulfilled for all $p \geq 2/3$. □

Lemma 5. *For $k = 2$ and $p \geq 2/3$, independence minimizes $\partial W/\partial p_1$.*

Proof. Independence decreases $\partial W/\partial p_1$ if:

$$p_A\frac{p}{3} - (1-p_A)\frac{(1-p)}{3} < p_A\frac{p^2}{2} - (1-p_A)\frac{(1-p)^2}{2}$$

which yields

$$p_A p \left(\frac{1}{3} - \frac{p}{2}\right) + (1-p_A)(1-p)\left(\frac{(1-p)}{2} - \frac{1}{3}\right) < 0$$

and is fulfilled for all $p \geq 2/3$. □

Proof $\partial W/\partial p_1 > 0$ for $p \geq 2/3$. Lemmas 4 and 5 show that $\beta(2) = 1$ and $\beta(0) = 0$ minimize $\partial W/\partial p_1 > 0$ for $p \geq 2/3$. Given this minimizing behavior of B, the derivative of the winning probability with respect to A_1's choice defined in Equation 11 becomes:

$$\begin{aligned}
\frac{\partial W}{\partial p_1} = & \; p_A p \frac{1}{3} \\
& + p_A(1-p)\frac{1}{3} \\
& + (1-p_A)p\frac{p}{2} \\
& + (1-p_A)(1-p)p \\
& - p_A p(1-p) \\
& - p_A(1-p)\frac{1-p}{2} \\
& - (1-p_A)p\frac{1}{3} \\
& - (1-p_A)(1-p)\frac{1}{3}
\end{aligned}$$

which boils down to

$$\frac{\partial W}{\partial p_1} = \frac{1}{6}\left(p(6-3p) - 2 - p_A(6p - 6p^2 - 1)\right)$$

which is greater or equal to zero for $0.5 \leq p$ and $p_A \leq p$. □

Equilibrium behavior of B

To complete the proof of Proposition 4, we specify the best response of B in equilibrium given that $q_1 = q_A = 1 \rightarrow p_1 = p_A = p$. For $k = 0$, B is conformist if $p > \frac{1}{2} + \frac{1}{2\sqrt{3}}$. For $k = 2$, B is anticonformist if $p < \frac{1}{6}\left(3 + \sqrt{6\sqrt{3} - 9}\right)$. This completes the proof of Proposition 4.

Proof. For $k = 0$, and $q_A = 1$, $p^H = 1 - p$, B follows the own taste if:

$$p^H p + (1-p^H)(1-p) \geq \frac{1}{3},$$

which yields

$$p + (1-p)(1-2p) \leq \frac{2}{3}, \tag{12}$$

and

$$p \leq \frac{1}{2} + \frac{1}{2\sqrt{3}} \approx 0.789 \tag{13}$$

For $k = 2$, and $q_A = 1$, $p^H = \frac{p^3}{3p^2-3p+1}$, B follows the own taste if:

$$\frac{1}{3} \geq p^H(1-p) + (1-p^H)p,$$

which yields

$$\frac{2}{3} \leq p + \frac{p^3(1-2p)}{3p^2-3p+1}, \tag{14}$$

and

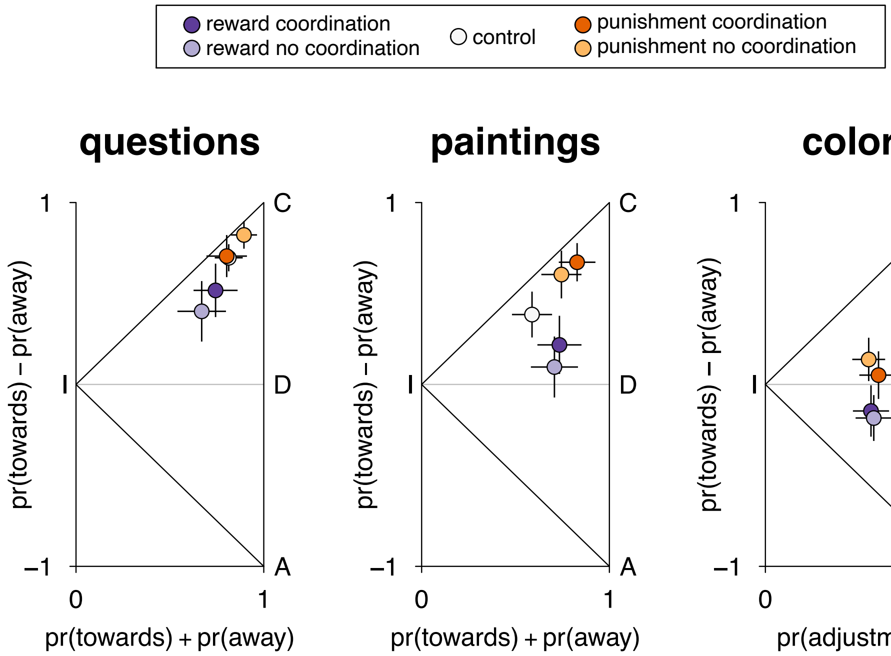
$$p \leq \frac{1}{6}\left(3 + \sqrt{6\sqrt{3}-9}\right) \approx 0.697 \tag{15}$$

□

B Additional results

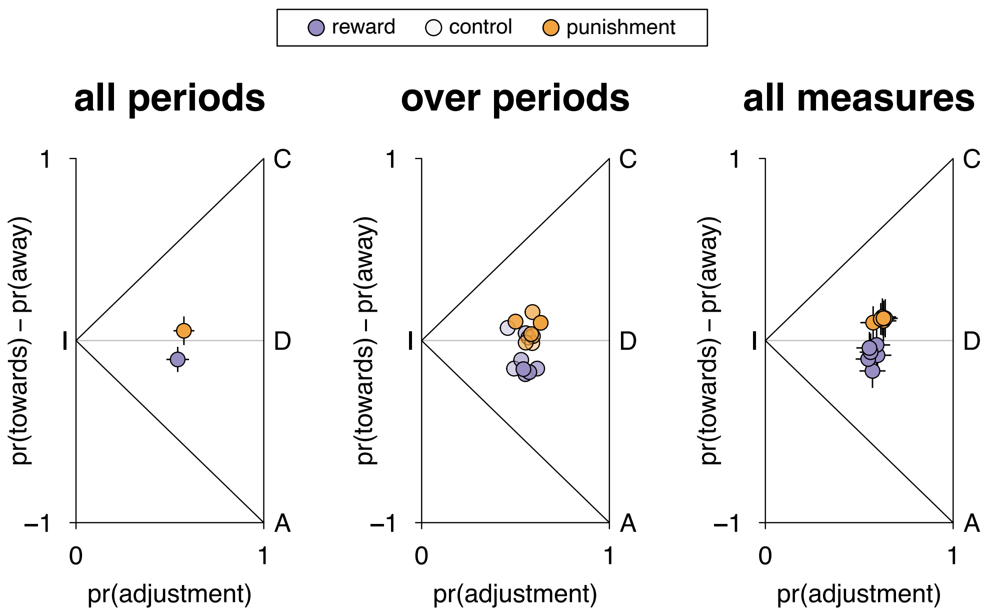
B.1 Supplementary figures and tables

Figure B.1: Average response to social influence across all treatments



Notes: Dots indicate the average response to social influence in a treatment. Whiskers indicate the 95 percent confidence interval based on the t-distribution and block-bootstrapped standard errors.

Figure B.2: Robustness of treatment effect in experiment two



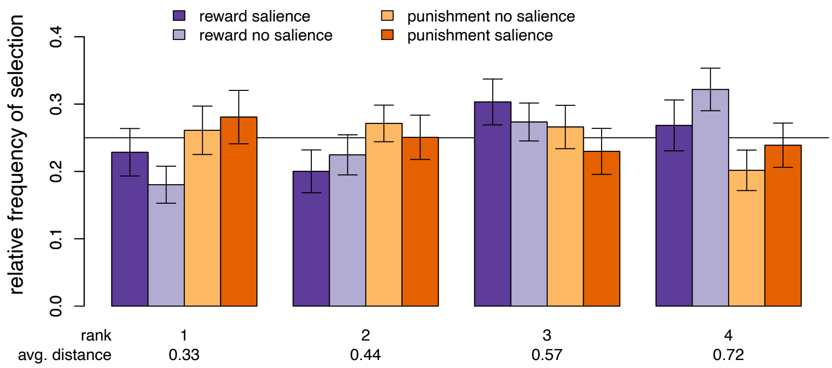
Notes: Left panel: average response to social influence for data of all periods. Central panel: evolution of the average response effect over periods. More intense colors reflect later periods. Right panel: average response for each of the 6 possible combinations of the three distance variables and the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

Table B.1: Average distance of adjusted choices across treatments

	reward	punishment	t-statistic	df	p-value
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
NO SALIENCE 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
SALIENCE 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

Notes: Table shows averages of matching group averages.

Figure B.3: Adjustment to distance ranks in experiment two



Notes: Relative frequency of distance ranks. Higher ranks indicate a larger minimum Euclidean distance of the selected color. Whiskers indicate 95% CIs, based on block-bootstrapped standard error of the mean (1000 samples, matching group ID).

B.2 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 by five-fold cross validation. The log likelihood of the mixture model is

$$\ln L = \sum_{i=1}^{N} \ln \left(\sum_{k=1}^{K} p_k \prod_{s=1}^{S} \prod_{r=1}^{R} (\pi_{ksr})^{y_{isr}} \right). \tag{16}$$

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.

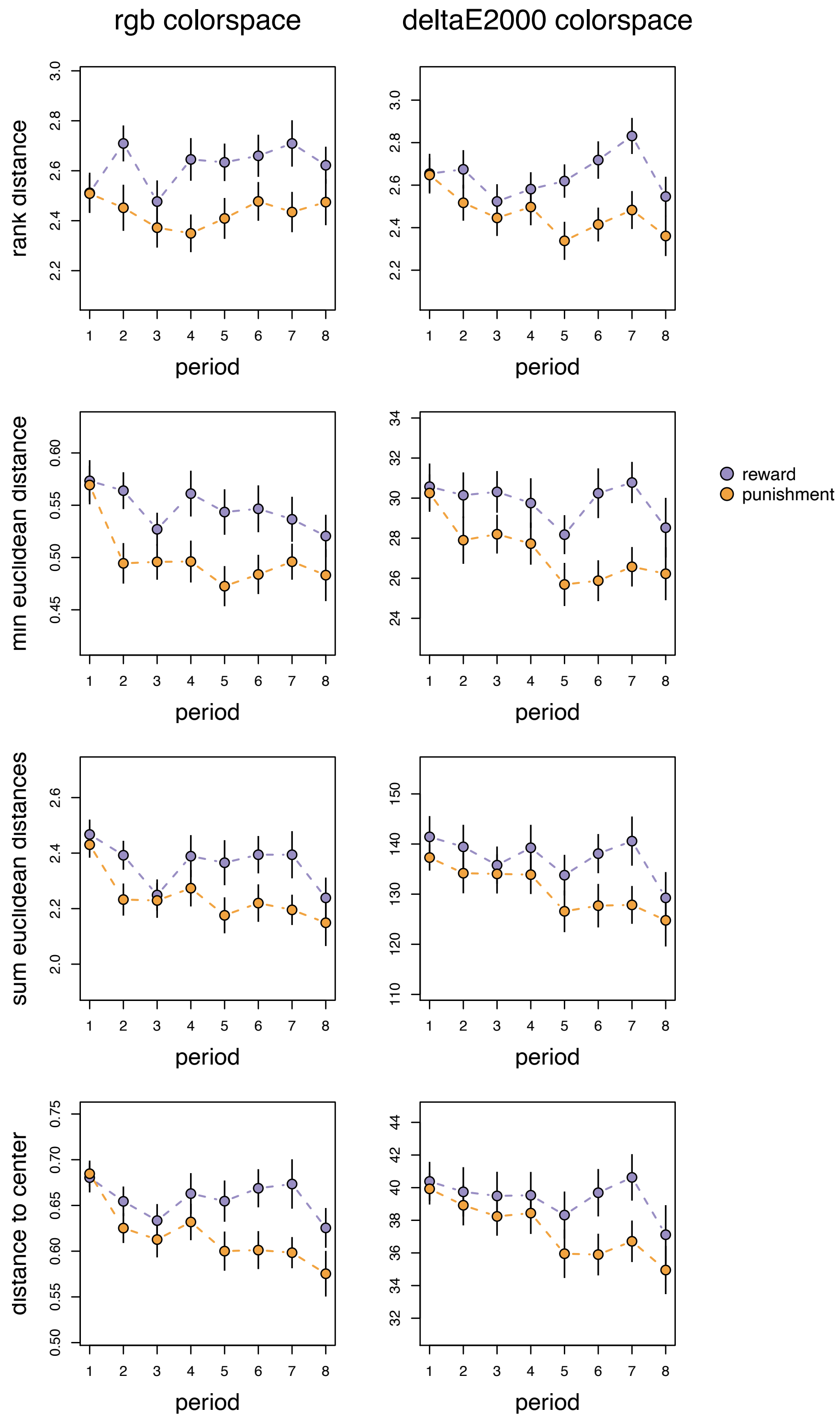
Estimates and standard errors of type position

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 one, $\pi_k^t = \pi_{ks\prime r\prime}$ 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 two, $\pi_k^t = \pi_{ksr\prime}$ where $r\prime$ indicates the response to adjust in the

Figure B.4: Evolution of distance measures over periods



Notes: Evolution of the average distance of social choices over the eight periods of experiment two. 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).

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.

Relating types to questionnaire data

We fit a latent class regression model to investigate whether the prior probability to belong to a certain type can be explained by the sum score of participants' answers in a post-experimental conformity questionnaire (Mehrabian and Stefl, 1995). The latent class regression model (Bandeen-Roche et al., 1997) assumes that the log odds of the prior probabilities p_{ik} that participant i is of type k is a linear function of covariates, as specifically, the k are modeled as

$$\ln(p_{ik}/p_{i1}) = X_i \beta_k \;\forall\; k \in K$$

where X is a $N \times 2$ matrix with an intercept in the first and the individual conformity scores in the second column, and β_k a column vector of coefficients. Algebraic manipulation of the K equations above yield

$$p_{ik} = \frac{e^{X_i \beta_k}}{\sum_{k=1}^{K} e^{X_i \beta_k}} \tag{17}$$

and the log-likelihood function of the latent class regression model is:

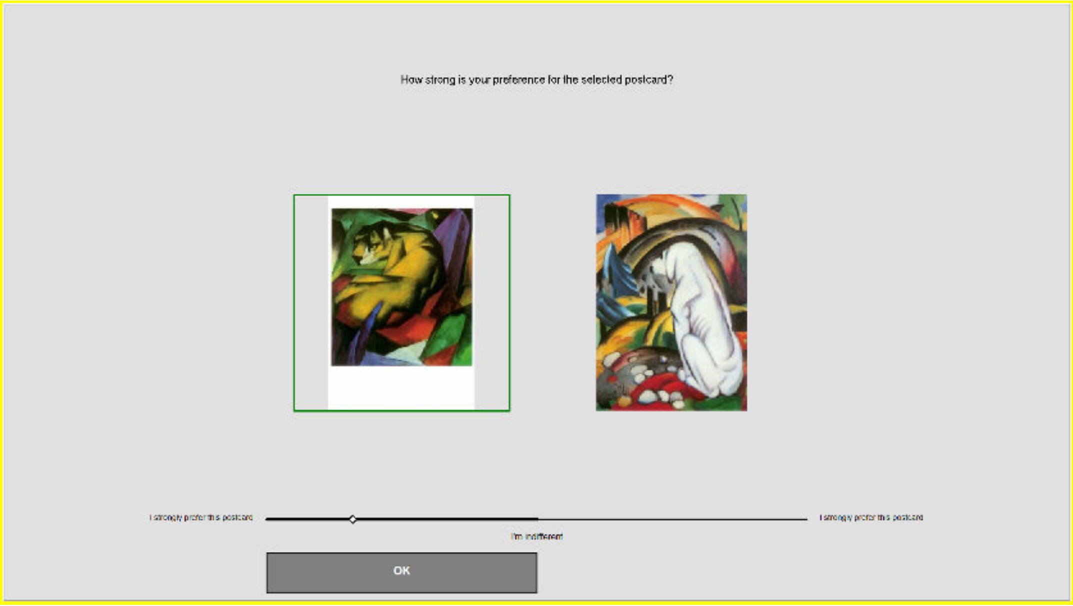
$$\ln L = \sum_{i=1}^{N} \ln \left(\sum_{k=1}^{K} p_{ik} \prod_{s=1}^{S} \prod_{r=1}^{R} (\pi_{ksr})^{y_{isr}} \right). \tag{18}$$

We use the R package *stratEst* (Dvorak, 2019) to obtain maximum likelihood estimates and block-bootstrapped standard errors of the parameters of the mixture and latent class regression models. The results are summarized in Table B.2.

C Study materials

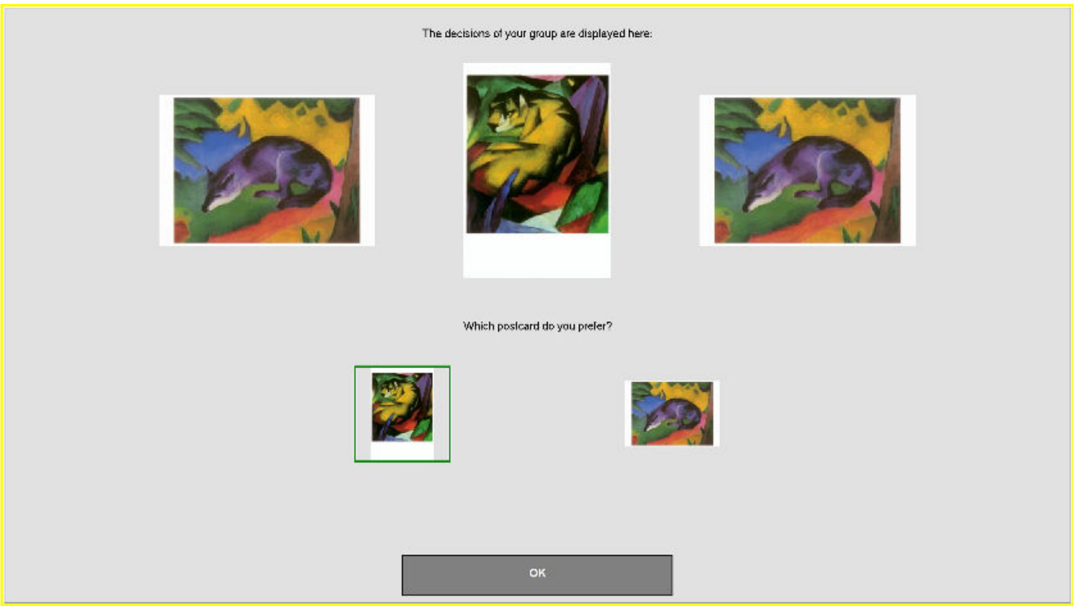
C.1 Decision screens

Figure C.5: Decision Screen of a related choice



Notes: Related choice comparing the one alternative of the social 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.

Figure C.6: Decision screen of a social choice



Notes: Social 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.

Table B.2: Mixture-Models for Social Response

	questions					paintings				colors		
	reward		control	punish		reward		control	punish	reward		punish
response type	1	2	1	1	2	1	2	1	1	1	2	1
share (%)	13	87	100	54	46	33	67	100	100	35	65	100
	(4)	(4)	-	(8)	(8)	(6)	(6)	-	-	(5)	(5)	-
π_k^t	0.11	0.67	0.75	0.95	0.65	0.18	0.57	0.49	0.71	0.25	0.18	0.34
	(0.09)	(0.04)	(0.03)	(0.04)	(0.06)	(0.06)	(0.06)	(0.05)	(0.04)	(0.10)	(0.3)	(0.03)
π_k^a	0.60	0.05	0.06	0.00	0.09	0.60	0.14	0.10	0.07	0.70	0.19	0.24
	(0.18)	(0.01)	(0.02)	-	(0.04)	(0.10)	(0.03)	(0.02)	(0.02)	(0.35)	(0.04)	(0.03)
latent class coefficients												
$\beta_{\text{intercept}}$	0	2.01	-	0	-0.18	0	0.83	-	-	0	0.62	
	-	(0.37)	-	-	(0.31)	-	(0.31)	-	-	-	(0.24)	
β_{score}	0	0.36	-	0	-0.04	0	0.87	-	-	0	0.35	
	-	(0.33)	-	-	(.30)	-	(0.32)	-	-	-	(0.23)	
N	105		60	105		105		60	105	155		166
Obs	571		356	607		569		360	590	312		345
LL	-255		-130	-179		-335		-171	-238	-318		-370

Notes: Maximum likelihood parameters of the fitted mixture models and standard errors in parentheses. Shares reflect the ML estimates of the frequency of a type in the treatment. The parameters π_k^t and π_k^a indicate the probability of adjustments in line with conformity and anticonformity. β_{score} is the coefficient for the sum score of the post-experimental conformity questionnaire (Mehrabian and Stefl, 1995). *N* indicates the number of participants, *Obs* the number of observations for each treatment. *LL* is the log likelihood of the model.

Table C.3: 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

Notes: 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 C.4: 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	9984670	9826675	9596960
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	1,861,484	1,904,569	1,964,375
2	Which country has more inhabitants (2014)?	France	Italy	UK	65,835,579	60,782,668	64,351,155
2	Which country has more inhabitants (2014)?	Spain	Ukraine	Poland	46,512,199	45,245,894	38,017,856
2	Which country has more inhabitants (2014)?	Greece	Belgium	Czech Republic	10,926,807	11,203,992	10,512,419
2	Which country has more inhabitants (2014)?	Austria	Switzerland	Bulgaria	8,506,889	8,139,631	7,245,677
2	Which country has more inhabitants (2014)?	Malta	Luxemburg	Iceland	425,384	549,680	325,671
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	Franklin 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. von 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	90,440,000	87,830,000	86,370,000
6	Which airline had more passengers?	Lufthansa	Easyjet	Air China	59,850,000	62,310,000	54,580,000
6	Which airline had more passengers?	Air Berlin	Brithish Airlines	Air France	29,910,000	41,160,000	45,410,000
6	Which airline had more passengers?	KLM	Aeroflot	SAS	27,740,000	23,600,000	27,390,000
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 in wealth (2012, GINI)?	France	Belgium	Austria	33.1	27.6	30.5
8	Which country has more inequality in wealth (2012, GINI)?	Norway	Finland	Sweden	25.9	27.1	27.3
8	Which country has more inequality in wealth (2012, GINI)?	Bolivia	Ecuador	Peru	46.7	46.6	45.1
8	Which country has more inequality in wealth (2012, GINI)?	Costa Rica	Brazil	Argentina	48.6	52.7	42.5
8	Which country has more inequality in wealth (2012, GINI)?	Thailand	Laos	Vietnam	39.3	37.9	38.7
9	Which soccer club has the higher market value (2016)?	Manchester City	FC Chelsea	Manchster United	501.75	490	411.25
9	Which soccer club has the higher market value (2016)?	AS Rom	FC Valencia	SSC Neapel	250.7	282	284
9	Which soccer club has the higher market value (2016)?	Bayer 04 Leverkusen	VfL Wolfsburg	FC Schalke 04	211.1	183.1	199.8
9	Which soccer club has the higher market value (2016)?	Zenit St.Petersburg	AC Mailand	FC Sevilla	198.6	188.1	186.2

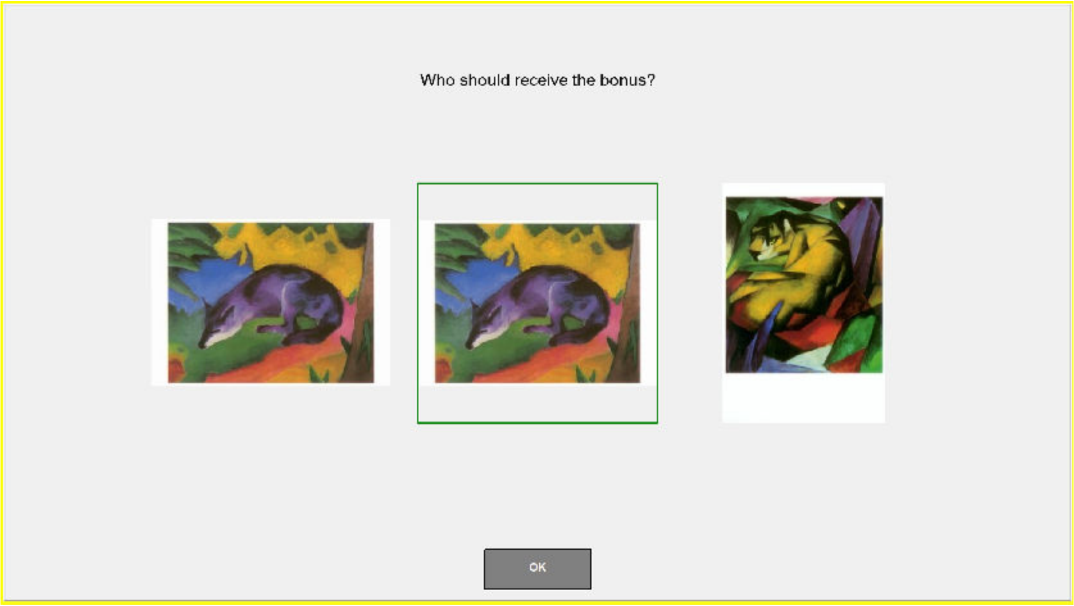
Sources: http://appsso.eurostat.ec.europa.eu, https://de.wikipedia.org, http://de.statista.com, http://carbonfootprintofnations.com, http://databank.worldbank.org, http://www.transfermarkt.de

Table C.5: 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 Winter Olympics?	Netherlands	France	Germany	24	15	19
10	Which country won more medals (2014 Winter Olympics?	Switzerland	Sweden	Austria	11	15	17
10	Which country won more medals (2014 Winter Olympics?	Canada	Norway	USA	26	28	
10	Which country won more medals (2014 Winter Olympics?	Finland	UK	Ukraine	5	4	2
10	Which country won more medals (2014 Winter Olympics?	Belarus	Kazakhstan	Australia	6	1	3
11	Which airport has been used by more passengers (2014)?	Atlanta Int	London Heathrow	Dubai Int	96,178,000	73,408,000	70,475,000
11	Which airport has been used by more passengers (2014)?	Singapore Changi	Kuala Lumpur Int	Shanghai Int	54,093,000	48,930,000	51,687,000
11	Which airport has been used by more passengers (2014)?	Paris Charles de Gaulles	Frankfurt am Main	Amsterdam Schiphol	63,813,000	59,566,000	54,978,000
11	Which airport has been used by more passengers (2014)?	Madrid Barajas	Sao Paulo-Guarulhos	Miami Int	41,822,000	39,765,000	40,941,000
12	Who sold more records in Germany?	The Beatles	Michael Jackson	Madonna	7,600,000	11,275,000	12,300,000
12	Who sold more records in Germany?	ACDC	ABBA	Robbie Williams	10,475,000	10,800,000	9,275,000
12	Who sold more records in Germany?	Helene Fischer	Pur	Die Aerzte	9,150,000	9,425,000	7,850,000
12	Who sold more records in Germany?	Britney Spears	Bon Jovi	Xavier Naidoo	5,050,000	5,150,000	5,525,000
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 people)?	USA	Cuba	Seychelles	698	510	799
15	Which country has more prisoners (2016, per 100k people)?	Thailand	Russia	Ruanda	468	447	434
15	Which country has more prisoners (2015, absolute number)?	Berlin	Saxony	Rhineland-Palatinate	3806	3385	3102
15	Which country has more prisoners (2015, absolute number)?	Saxony-Anhalt	Thuringia	Hamburg	1670	1600	1559
15	Which country has more prisoners (2015, absolute number)?	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 share)?	Guinea	Niger	Burkina Faso	74.7	84.5	71.3
19	Which country has more analphabets (relative share)?	Mali	Chad	Ethiopia	66.4	62.7	61
19	Which country has more analphabets (relative share)?	Liberia	Haiti	Sierra Leone	57.1	51.3	55.5
19	Which country has more analphabets (relative share)?	Pakistan	Bhutan	Senegal	45.3	47.2	47.9
19	Which country has more analphabets (relative share)?	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

Figure C.7: Decision screen of evaluator



Notes: 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.

C.2 Instructions of experiment one

Instructions

We ask you to keep quiet at your computer workstation and not to communicate with others during the experiment. Anyone who intentionally violates this rule is requested to leave the experiment without payment. If you have any questions, please contact us and wait until an experimenter comes to you. Income is calculated in points. At the end of the experiment, the total amount of points you have earned will be converted into euros at the following rate:

1 point = 1 Euro

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

Please read the instructions carefully. When all the instructions have been read, answer some comprehension questions. Then make your decisions in the experiment. Your decisions will be treated anonymously.

General procedure

This experiment consists of two parts, each consisting of three stages. In each step you make several decisions. Your total income is the sum of your income from both parts. At the beginning of the first part, you are 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 have no effect on your income in the second part.

Which postcard do you choose?

Overview

In this part you decide between two art postcards. The motives of the two cards are displayed on the screen and you choose which of the two motives you prefer. After all group members have made their decisions, the motifs selected by the three persons are shown to judges. Based on the selected motives, each reviewer marks one person in 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, a decision situation for your group is randomly selected. You will receive your preferred motif from this situation as a real postcard. A different pair of postcards will be drawn for each group, from which the group members will receive their preferred card. It is therefore only possible for members of your group to receive the same postcard as you did at the end of the experiment. Your score will be credited 10 points for editing this part. If you are the marked person in the selected decision situation, a further 10 points will be added to your score. 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.

This part consists of three levels, which you process one after the other. The levels are described in more detail below.

Stage 1 You will see two postcard images on the screen, as shown in Figure 1. You decide which postcard you prefer. To do this, click on the corresponding motif. After each decision, we will ask you to indicate your preference for the motif you have chosen. After your decision, the bar shown in figure 1 will appear below the pictures. You make these decisions one after the other for 20 pairs of postcards. The members of a group may receive different pairs to choose from.

Stage 2 At this stage, you can also choose one of two postcard motifs. The other two members of your group have already passed through the decision situations that you process in step 2 in step 1. Before each decision, you can see how your group members have decided on the respective pair of postcards (upper part in Figure 2a). You now select a motif again and indicate how much you like the motif (lower part in Figure 2b). You make this decision one after the other for 10 pairs of postcards. With your decision you will see in the upper row the 3 postcards 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, with the order of the 3 postcards on the evaluators' screens being random and different for each evaluator.

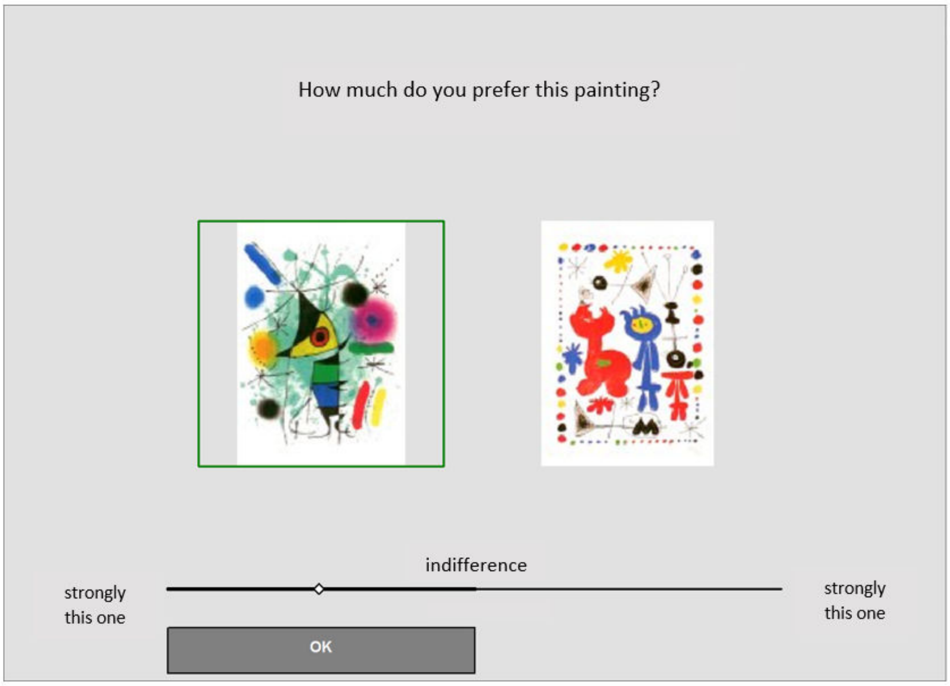


Figure 1

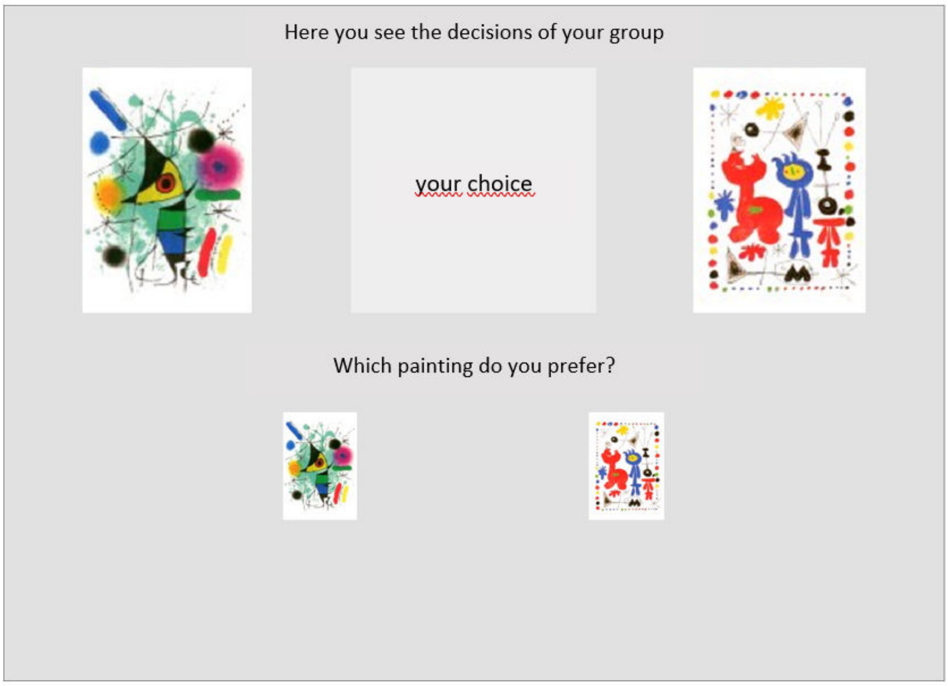


Figure 2a

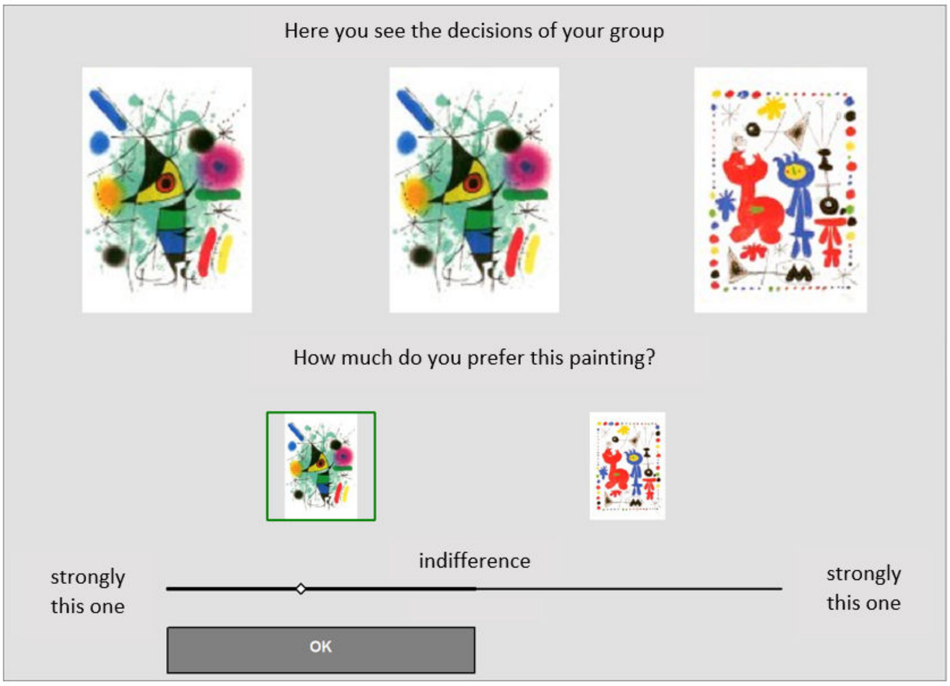


Figure 2b

Stage 3 For a decision situation the 3 selected motives of your group are transmitted to members of other groups for evaluation. Each evaluator should mark a person in your group on the basis of the selected motives, who may then receive a bonus of 10 points. The payout of an appraiser is all the higher the more of the other appraisers mark the same person as he/she. You also decide as an appraiser. For a decision situation, you can see how the three members of another group have decided and mark who should receive the bonus on the basis of the selected motives. The more of the other appraisers decide the same as you do, the higher your payout as an appraiser. At the time of its decision one does not know as an appraiser however yet, which decision situation in a group is drawn by lots for the payment and how the three group members actually decided in this situation. You therefore specify who is to 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. So the selected motives of a person appear sometimes left, sometimes in the middle and sometimes right, and the positions are shuffled for each decision and each reviewer. At the end, the decision situation and appraiser relevant to your group will be drawn by lot. One member per group receives the bonus. Example: Below you can see various situations that can arise in a group when you choose 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 chose another. You select who is to receive the bonus as the appraiser. To do this, click on the relevant motif. Your selection is marked by a green frame.

Ultimately, the constellation that actually occurred in the group assigned to you at random always applies. For example, if the group members have decided as shown in Figure 3a and all selected the same theme, the person you selected for this

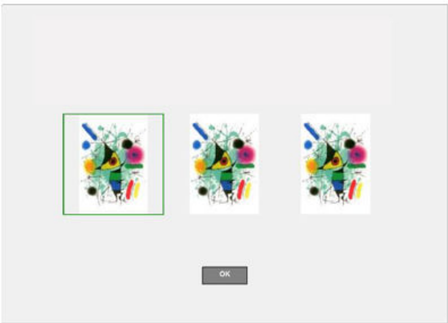


Figure 3a

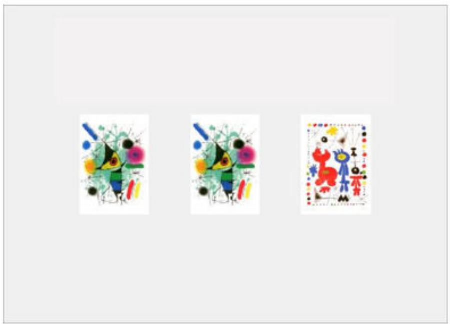


Figure 3b

constellation receives the bonus if applicable. As with all decisions, the appraisers' decisions are mutually anonymous. Neither the selected person nor the appraiser 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 appraisers receive additional payments for their decisions. These payouts are higher the more similarities you have with other appraisers. In concrete terms, you (and all other appraisers) receive 0.02 points per 10% match for each situation. Therefore, if in one situation 10% of the other appraisers selected the same participant as you, you receive 0.02 points; if you match half (50%) of the other appraisers, you receive 0.1 points; and if you match all other appraisers (100%), you receive 0.2 points. According to this principle, your payout is calculated and added up for each assessment situation. Note that the order of participants displayed is random and may be different for each assessor.

graduation Finally, one decision situation per group is selected at random. The motives 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 at home. You will then know which postcard you will receive based on your decision in the randomly drawn decision situation. You will also find out whether you were the marked person in this decision situation and thus receive a bonus. For each of your decisions as an evaluator, you will find out to what extent you agree with other evaluators and what payment you receive in return. You will receive your postcard at the end of the experiment together with the payment. If you have any questions, you can contact us 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 you don't miss out when the experiment continues.

Which answer do you choose?

For this part you will be divided into a new group of three. In the first part, the members of your new group were divided into three different groups. All processes in this part are the same as in the previous part - with one difference: You don't decide between art postcards, but between two answers to a factual question. The question and the two answers are displayed on the screen and you select one of the answers. After all group members have made their decisions, the answers selected by the three people are shown to judges from other groups. Based on the answers selected, each reviewer marks one person in your group, who may then receive a bonus. The order of the 3 answers on a reviewer's screen is random again and is shuffled in each decision situation and for each reviewer. Also you decide as an evaluators for other groups. All judges receive again an additional payment, which is all the higher, the more frequently one agrees with other judges. These decisions are met for several fact questions. At the end of the experiment a decision situation for your group is selected coincidentally. You will then know whether your answer was correct in this situation. For each group a different factual question is drawn. Your score will be credited 10 points for working on this part. If you are the person marked by the selected evaluator in the selected decision situation, a further 10 points will be added to your score. You will again receive 0.02 points for your decisions as an assessor for every 10This part also consists of the three levels that you have already worked on one after the other in the last 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.3 Instructions of experiment two

Instructions

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

We ask you to keep quiet at your computer and not to communicate with the other participants during the experiment. Your mobile phones should now be switched off. If you carry a switched on device with you, please switch it off immediately and place it in the designated rack. Anyone who violates these rules will be asked to leave the experiment without compensation. Your income 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 exchange rate:

1 point = 1 Euro

You will then receive your total income in cash. Your decisions and payouts 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 part when the first part is completed. Your decisions in the first part do not affect your income in the following parts.

Part 1

Division into groups

Before the beginning of the experiment you will be divided into groups of 6 persons. In part 1 you will only interact with participants of your own group. There are two roles, designer and evaluator. Each group consists of 4 designers and 2 evaluators. You will learn your role shortly before part 1 begins. The assigned roles remain the same throughout the experiment. At the beginning each designer receives a basic equipment of 12 points. Each evaluator also receives a basic equipment of 12 points at the beginning.

General procedure

Part 1 consists of 8 rounds. Each round has the same sequence and consists of four phases: Design phase, publication phase, evaluation phase and feedback phase. During the design phase, each designer generates several colours. In the publication phase, exactly one colour is published by each designer and shown to the judges. In the evaluation phase, each reviewer then selects a designer based on the four published colours. If both reviewers choose the same designer's color, they will receive an additional payout. In addition, an evaluator is drawn by lot whose decision determines which of the designers receives a bonus of 2 points. In the feedback phase, the designers are informed who received the bonus.

I Design phase

In the design phase, each designer generates colors by mixing. All designers have 2 minutes in each round to do this. During this time, the judges 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: Work area, selection area and history.

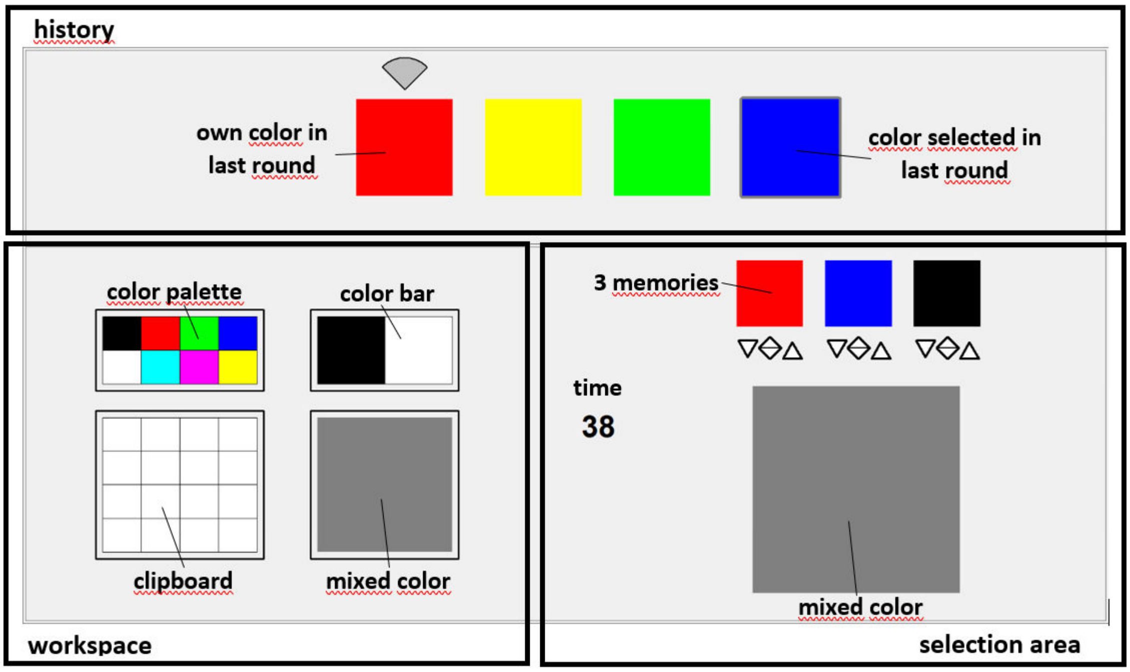


Figure 1: Screen for color generation

Workspace: New colors are generated in the lower left workspace. To do this, 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 with each other and the result appears directly below as a mixed color. In order to further process mixed colors, they can first be dragged to the clipboard with the left mouse button pressed down and then used again for mixing.

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

History: In the history, starting with round 2, the designers see all colors published by the designers in their group in the previous round. Their own color is indicated by a symbol, the color selected for the bonus is bordered in gray.

II Release phase

In each round, a color is published by each designer. Only colors that are in their own 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 must first make a preselection and then a conditional selection.

Preselection: When the time of the design phase has expired, each designer first makes a preselection. The designer selects one of the four colors in his or her selection area. The colours pre-selected by the designers are then temporarily displayed in the history to all other designers in their own group. The judges do not see the pre-selected colors.

Conditional selection: In the conditional selection that follows, each designer can adjust his or her decision based on the results of the preselection. To do this, the designer again selects one of his four colors. Transmission: The colour of the preselection is now transmitted to the judges by three designers. However, a randomly selected designer in each round will submit the conditional selection, which he has determined based on the three pre-selected colors submitted. All non-selected colors of the designers remain private. This means that no other participant will see them at any time during this experiment.

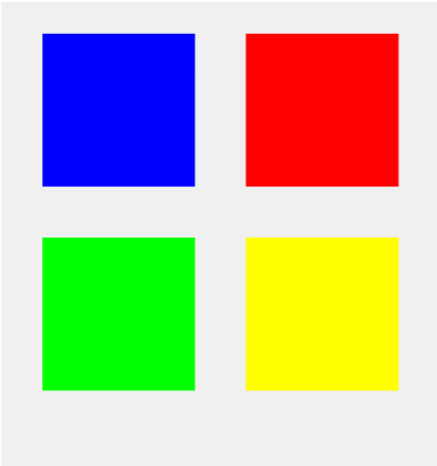
III Evaluation Phase

The four transmitted colors of the designers of a group are now displayed to the evaluators (Figure 2a). The arrangement of the colours in each round and for each assessor is determined randomly. The position of a designer's colours therefore changes both over the rounds and across the appraisers. Thus on the one hand from the position no conclusions on preceding publications can be closed. On the other hand it is probable that with the two judges at the same position different colors are indicated. Each assessor selects now by mouse-click one of the colors. The judges do not know thereby whether a color comes from the preselection or the conditioned selection. The two appraisers receive an additional payment of 2 points if both have chosen the same designer on the basis of the color. In the example in Figure 2b, the relevant appraiser has selected the yellow color (indicated by the gray border). Only if the irrelevant appraiser also chose yellow do both appraisers receive an additional 2 points. If the appraisers have chosen different colors, they receive no additional payment. This does not change the designer's score.

Before the first round begins, one of the two appraisers whose decision is relevant for the designers is drawn by lot. This relevant appraiser is the same person in all rounds. The other appraiser is irrelevant for the designers. However, the appraisers themselves do not know which of them is the relevant appraiser. The designer of the color selected by the relevant appraiser receives a bonus of 2 points. This does not change the appraisers' score.

IV Feedback phase

Once the evaluators have decided, the designers see the relevant evaluator's decision (Figure 2b, bordered in grey). The designers do not see the decision of the irrelevant evaluator. A designer's own color is highlighted with a white symbol. If several identical colors have been submitted, each designer can see in this way whether he or she has received the bonus. At the end of a round, the appraisers do



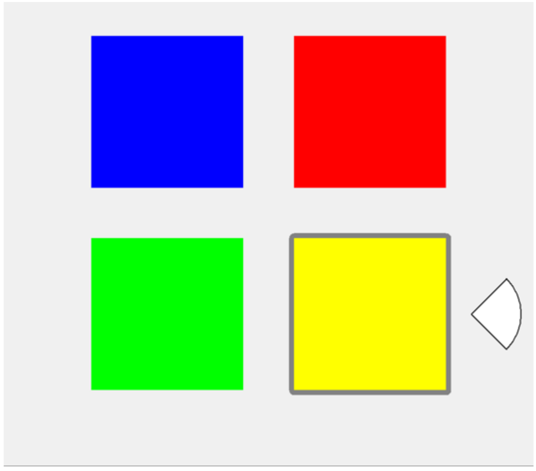


Figure 2a: evaluator selection screen **Figure 2b: designer feedback screen**

not yet know how the other appraiser decided. 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 it.

If you have any further questions, please contact us by hand signal. If you have no more questions and are ready for part 1, please click on "Experiment" in the upper right corner and then on "Inform".

References

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

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

Allen, V. L. (1965): "Situational Factors In Conformity," *Advances in Experimental Social Psychology*, 2, 133 – 175.

Allport, F. H. (1934): "The J-Curve Hypothesis of Conforming Behavior," *The Journal of Social Psychology*, 5, 141–183.

Alpizar, F., F. Carlsson, and O. Johansson-Stenman (2008): "Anonymity, reciprocity, and conformity: Evidence from voluntary contributions to a national park in Costa Rica," *Journal of Public Economics*, 92, 1047 – 1060.

Amabile, T. M., P. Goldfarb, and S. C. Brackfleld (1990): "Social influences on creativity: Evaluation, coaction, and surveillance," *Creativity Research Journal*, 3, 6–21.

Ambady, N. and R. Rosenthal (1992): "Thin slices of expressive behavior as predictors of interpersonal consequences: A meta-analysis." *Psychological Bulletin*, 111, 256 – 274.

Anderson, L. R. and C. A. Holt (1997): "Information Cascades in the Laboratory," *American Economic Review*, 87, 847–862.

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

Ariely, D. and J. Levav (2000): "Sequential Choice in Group Settings: Taking the Road Less Traveled and Less Enjoyed," *Journal of Consumer Research*, 27, 279–290.

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

——— (1952): *Social Psychology*, New-Jersey: Prentice-Hall.

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

——— (1956): "Studies of independence and conformity: A minority of one against a unanimous majority," *Psychological Monographs*, 70.

Bandeen-Roche, K., D. L. Miglioretti, S. L. Zeger, and P. J. Rathouz (1997): "Latent Variable Regression for Multiple Discrete Outcomes," *Journal of the American Statistical Association*, 92, 1375–1386.

Bandura, A. and R. H. Walters (1977): *Social learning theory*, vol. 1, Prentice-hall Englewood Cliffs, NJ.

Banerjee, A. and D. Fudenberg (2004): "Word-of-mouth learning," *Games and Economic Behavior*, 46, 1 – 22.

Banerjee, A. V. (1992): "A Simple Model of Herd Behavior," *The Quarterly Journal of Economics*, 107, 797–817.

Baron, R. S., J. A. Vandello, and B. Brunsman (1996): "The forgotten variable in conformity research: Impact of task importance on social influence." *Journal of personality and social psychology*, 71, 915.

Becker, G. (1971): *The Economics of Discrimination*, University of Chicago Press, 2 ed.

Bernheim, B. D. (1994): "A Theory of Conformity," *Journal of Political Economy*, 102, 841–877.

Bernheim, B. D. and C. Exley (2015): "Understanding Conformity: An Experimental Investigation," Tech. rep., Harvard Business School NOM Unit Working Paper No. 16-070. Available at SSRN: https://ssrn.com/abstract=2703932 or http://dx.doi.org/10.2139/ssrn.2703932.

Bertrand, M. and E. Duflo (2017): "Chapter 8 - Field Experiments on DiscriminationaaLaura Stilwell and Jan Zilinsky provided excellent research assistance. We thank Abhijit Banerjee for comments. We are particularly grateful to Betsy Levy Paluck, our discussant, for her detailed and thoughtful review of an earlier draft." in *Handbook of Field Experiments*, ed. by A. V. Banerjee and E. Duflo, North-Holland, vol. 1 of *Handbook of Economic Field Experiments*, 309 – 393.

Bikhchandani, S., D. Hirshleifer, and I. Welch (1992): "A Theory of Fads, Fashion, Custom, and Cultural Change as Informational Cascades," *Journal of Political Economy*, 100, 992–1026.

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, 49–64.

Bond, M. H. and P. B. Smith (1996): "Cross-Cultural Social and Organizational Psychology," *Annual Review of Psychology*, 47, 205–235, pMID: 15012481.

Boyd, R. and P. J. Richerson (1982): "Cultural transmission and the evolution of cooperative behavior," *Human Ecology*, 10, 325–351.

——— (1985): *Culture and the evolutionary process*, Chicago: University of Chicago Press.

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

Cialdini, R. B. and N. J. Goldstein (2004): "Social Influence: Compliance and Conformity," *Annual Review of Psychology*, 55, 591–621, pMID: 14744228.

Cialdini, R. B., W. Wosinska, D. W. Barrett, J. Butner, and M. 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, 1242–1253.

Clark, P. J. and F. C. Evans (1954): "Distance to Nearest Neighbor as a Measure of Spatial Relationships in Populations," *Ecology*, 35, 445–453.

Coleman, S. (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, L. and B. Greiner (2007): "Herding, social preferences and (non-)conformity," *Economics Letters*, 97, 74 – 80.

Croissant, Y. (2019): *mlogit: Multinomial Logit Models*, r package version 0.4-1.

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

Cunningham, W. A., M. K. Johnson, C. L. Raye, J. C. Gatenby, J. C. Gore, and M. R. Banaji (2004): "Separable Neural Components in the Processing of Black and White Faces," *Psychological Science*, 15, 806–813, pMID: 15563325.

Dvorak, F. (2019): *stratEst: Strategy Estimation in R*, r package version 0.1.2.

Efferson, C., R. Lalive, P. J. Richerson, R. McElreath, and M. Lubell (2008): "Conformists and mavericks: the empirics of frequency-dependent cultural transmission," *Evolution and Human Behavior*, 29, 56–64.

Falk, A. and F. Zimmermann (2017): "Consistency as a Signal of Skills," *Management Science*, 63, 2197–2210.

Fatas, E., S. P. H. Heap, and D. R. Arjona (2018): "Preference conformism: An experiment," *European Economic Review*, 105, 71 – 82.

Fehr, E. and U. Fischbacher (2004): "Social norms and human cooperation," *Trends in Cognitive Sciences*, 8, 185 – 190.

Fehr, E., U. Fischbacher, and S. Gächter (2002): "Strong reciprocity, human cooperation, and the enforcement of social norms," *Human Nature*, 13, 1–25.

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

Fischbacher, U. (2007): "Z-Tree: Zurich toolbox for ready-made economic experiments," *Experimental Economics*, 10, 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, 521–529.

Gama, J. and G. Davis (2018): *colorscience: Color Science Methods and Data*, r package version 1.0.5.

Goeree, J. K. and L. Yariv (2015): "Conformity in the lab," *Journal of the Economic Science Association*, 1, 15–28.

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

Griskevicius, V., N. J. Goldstein, C. R. Mortensen, R. B. Cialdini, and D. T. Kenrick (2006): "Going Along Versus Going Alone: When Fundamental Motives Facilitate Strategic (Non)Conformity," *Journal of Personality and Social Psychology*, 91, 281–294.

Gürerk, Ö., B. Irlenbusch, and B. Rockenbach (2006): "The Competitive Advantage of Sanctioning Institutions," *Science*, 312, 108–111.

Guzmán, R. A., C. Rodriguez-Sickert, and R. Rowthorn (2007): "When in Rome, do as the Romans do: the coevolution of altruistic punishment, conformist learning, and cooperation," *Evolution and Human Behavior*, 28, 112 – 117.

Hamilton, W. (1971): "Geometry for the selfish herd," *Journal of Theoretical Biology*, 31, 295 – 311.

Henrich, J., R. McElreath, A. Barr, J. Ensminger, C. Barrett, A. Bolyanatz, J. C. Cardenas, M. Gurven, E. Gwako, N. Henrich, C. Lesorogol, F. Marlowe, D. Tracer, and J. Ziker (2006): "Costly Punishment Across Human Societies," *Science*, 312, 1767–1770.

Hewstone, M., M. Rubin, and H. Willis (2002): "Intergroup Bias," *Annual Review of Psychology*, 53, 575–604, pMID: 11752497.

Huck, S. and J. Oechssler (2000): "Informational cascades in the laboratory: Do they occur for the right reasons?" *Journal of Economic Psychology*, 21, 661 – 671.

Imhoff, R. and H.-P. Erb (2009): "What Motivates Nonconformity? Uniqueness Seeking Blocks Majority Influence," *Personality and Social Psychology Bulletin*, 35, 309–320, pMID: 19098256.

Jones, D. and S. Linardi (2014): "Wallflowers: Experimental Evidence of an Aversion to Standing Out," *Management Science*, 60, 1757–1771.

Jones, S. R. G. (1984): *The Economics of Conformism*, Oxford: Blackwell.

Juul, J. S. and M. A. Porter (2019): "Hipsters on networks: How a minority group of individuals can lead to an antiestablishment majority," *Physical Review E*, 99.

Kameda, T., R. S. Tindale, and J. H. Davis (2003): "Cognitions, preferences, and social sharedness: Past, present, and future directions in group decision making," *Emerging perspectives on judgment and decision research*, 458.

Kelman, H. C. (1961): "Processes of Opinion Change," *Public Opinion Quarterly*, 25, 57–78.

Kets, W. and A. 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.

Kim, H. and H. R. Markus (1999): "Deviance or uniqueness, harmony or conformity? A cultural analysis." *Journal of Personality and Social Psychology*, 77, 785–800.

Krupka, E. L. and R. A. Weber (2013): "Identifying Social Norms Using Coordination Games: Why Does Dictator Game Sharing Vary?" *Journal of the European Economic Association*, 11, 495–524.

Lee, I. H. (1993): "On the Convergence of Informational Cascades," *Journal of Economic Theory*, 61, 395 – 411.

——— (1998): "Market Crashes and Informational Avalanches," *The Review of Economic Studies*, 65, 741–759.

Lynn, M. and J. Harris (1997): "Individual Differences in the Pursuit of Self-Uniqueness Through Consumption," *Journal of Applied Social Psychology*, 27, 1861–1883.

March, J. G. (1991): "Exploration and Exploitation in Organizational Learning," *Organization Science*, 2, 71–87.

McElreath, R., A. V. Bell, C. Efferson, M. Lubell, P. J. Richerson, and T. 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, 3515–3528.

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

Mehrabian, A. and C. A. Stefl (1995): "Basic Temperament Components of Loneliness, Shyness, and Conformity," *Social Behavior and Personality*, 23, 253–263.

Mesoudi, A., L. Chang, S. R. Dall, and A. Thornton (2016): "The evolution of individual and cultural variation in social learning," *Trends in ecology & evolution*, 31, 215–225.

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

Nail, P. R., S. I. Di Domenico, and G. MacDonald (2013): "Proposal of a Double Diamond Model of Social Response," *Review of General Psychology*, 17, 1–19.

Nail, P. R. and M. D. Van Leeuwen (1993): "An Analysis and Restructuring of the Diamond Model of Social Response," *Pers Soc Psychol Bull*, 19, 106–116.

Nolan, J. M., P. W. Schultz, R. B. Cialdini, N. J. Goldstein, and V. Griskevicius (2008): "Normative Social Influence is Underdetected," *Personality and Social Psychology Bulletin*, 34, 913–923, pMID: 18550863.

Nöth, M. and M. Weber (2002): "Information Aggregation with Random Ordering: Cascades and Overconfidence," *The Economic Journal*, 113, 166–189.

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

Nyczka, P. and K. Sznajd-Weron (2013): "Anticonformity or Independence? Insights from Statistical Physics," *Journal of Statistical Physics*, 151, 174–202.

Riach, P. A. and J. Rich (2002): "Field Experiments of Discrimination in the Market Place," *The Economic Journal*, 112, F480–F518.

Robin, S., A. Rusinowska, and M. C. Villeval (2014): "Ingratiation: Experimental evidence," *European Economic Review*, 66, 16 – 38.

Rubinstein, A. (2013): "10 Q&A: Experienced Advice forp "Lost" Graduate Students in Economics," *The Journal of Economic Education*, 44, 193–196.

Sakha, S. and A. 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, T. M., P. W. Dillingham, and A. J. W. Ward (2017): "The effects of external cues on individual and collective behavior of shoaling fish," *Science Advances*, 3.

Schultz, P. W., J. M. Nolan, R. B. Cialdini, N. J. Goldstein, and V. Griskevicius (2007): "The Constructive, Destructive, and Reconstructive Power of Social Norms," *Psychological Science*, 18, 429–434.

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

Shalley, C. E. and J. 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 – 22.

SHERIF, M. (1935): "A study of some social factors in perception," *Archives of Psychology*, 187.

SIEDLECKI, P., J. SZWABINSKI, AND W. TOMASZ (2016): "The Interplay Between Conformity and Anticonformity and its Polarizing Effect on Society," *Journal of Artificial Societies and Social Simulation*, 19.

SMITH, L. AND P. SORENSEN (2000): "Pathological Outcomes of Observational Learning," *Econometrica*, 68, 371–398.

SMITH, S., F. WINDMEIJER, AND E. WRIGHT (2015): "Peer Effects in Charitable Giving: Evidence from the (Running) Field," *The Economic Journal*, 125, 1053–1071.

SOSNA, M. M. G., C. R. TWOMEY, J. BAK-COLEMAN, W. POEL, B. C. DANIELS, P. ROMANCZUK, AND I. D. COUZIN (2019): "Individual and collective encoding of risk in animal groups," *Proceedings of the National Academy of Sciences*, 116, 20556–20561.

TOUBOUL, J. (2019): "The hipster effect: When anti-conformists all look the same," *Discrete & Continuous Dynamical Systems - B*, 24, 4379–4415.

VINE, I. (1971): "Risk of visual detection and pursuit by a predator and the selective advantage of flocking behaviour," *Journal of Theoretical Biology*, 30, 405 – 422.

VISCIDO, S. V. AND D. S. WETHEY (2002): "Quantitative analysis of fiddler crab flock movement: evidence for selfish herd behaviour," *Animal Behaviour*, 63, 735 – 741.

VIVES, X. (1997): "Learning from Others: A Welfare Analysis," *Games and Economic Behavior*, 20, 177 – 200.

WILLIS, R. H. (1963): "Two Dimensions of Conformity-Nonconformity," *Sociometry*, 26, 499–513.

——— (1965): "Conformity, Independence, and Anticonformity," *Human Relations*, 18, 373–388.

WILLIS, R. H. AND J. M. LEVINE (1976): *Social psychology: An introduction*, New York: Free Press, chap. Interpersonal influence and conformity, 309–341.

WINTER, L. AND J. S. ULEMAN (1986): "'When are social judgments made? Evidence for the spontaneousness of trait inferences': Correction to Winter and Uleman." *Journal of Personality and Social Psychology*, 50, 355.

WORCHEL, S. AND J. W. BREHM (1970): "Effect of threats to attitudinal freedom as a function of agreement with the communicator." *Journal of Personality and Social Psychology*, 14, 18–22.

WRIGHT, D. B., K. LONDON, AND M. WAECHTER (2009): "Social anxiety moderates memory conformity in adolescents," *Applied Cognitive Psychology*, 24, 1034–1045.

Yamagishi, T., H. Hashimoto, and J. Schug (2008): "Preferences Versus Strategies as Explanations for Culture-Specific Behavior," *Psychological Science*, 19, 579–584, pMID: 18578848.

THURGAU INSTITUTE
OF ECONOMICS
at the University of Konstanz

Hauptstrasse 90
Postfach
CH-8280 Kreuzlingen 2

T +41 (0)71 677 05 10
F +41 (0)71 677 05 11

info@twi-kreuzlingen.ch
www.twi-kreuzlingen.ch