Essays on Status and Conflict Behavior

Inaugural-Dissertation

zur Erlangung des Grades

Doctor oeconomiae publicae (Dr. oec. publ.)

an der Ludwig-Maximilians-Universität München

2022

vorgelegt von

Biljana Meiske

Referent: Prof. Dr. Kai A. Konrad

Korreferent: Prof. Dr. Davide Cantoni

Promotionsabschlussberatung: 20 Juli 2022

Contents

List of Figures

List of Tables

Introduction

Throughout history, humans formed groups based on a variety of inclusion criteria in order to structure their interactions, including both cooperative and conflict behavior. Whereas humans exhibit levels of non-kinship cooperation that still require a definite theoretical explanation (Egas et al. (2013), Nowak (2006), Fehr and Fischbacher (2003)), at the same time, the conflict between individuals and groups is pervasive, often entailing high socio-economic costs for the involved parties (Cikara and Van Bavel (2014), Cohen and Insko (2008), Choi and Bowles (2007)). Warfare, including inter-state wars and civil conflicts, stands as a clear example. Other phenomena, such as competition for political dominance, rent-seeking by interest groups, or organized crime gangs' competition for market dominance present other examples of more or less violent forms of conflict over valuable resources (Hirshleifer (1989), Wärneryd (2014)), which all can have far-reaching consequences for the economic functioning of the environments where they take place.

Besides directly contesting the control of scarce economic resources, conflicts also arise over more abstract concepts such as social status. Indeed, social status, as "the relative rank of an individual along one or more social dimensions within a given social hierarchy" (Mattan et al. (2017)), has been found to profoundly affect relationships within and between groups. Social status typically caries with it the expectation of entitlement to certain resources (Ball et al. (2001)) and the high status individuals have been found to receive greater attention (Dalmaso et al. (2012), Foulsham et al. (2010)) and help from the lower status individuals (Van der Vegt et al. (2006)), are accorded higher influence (Ridgeway and Walker (1995)), and are seen as more desirable teammates (Hardy and Van Vugt (2006)). Insofar, either as an intermediate good or a good perceived to have value on its own, social status has far-reaching consequences for those who hold it and

can serve as a powerful motivator of both conflict and cooperative behavior. In the three chapters of this thesis, I study selected aspects of conflict behavior and implications of status concerns and attempt to provide regularities behind them.

The first chapter focuses on analyzing conflict behavior in the aftermath of productivity shocks. I start off from the observation of the profound effect of productivity and productivity shocks on the probability of conflict outbreaks and intensity established in the empirical literature on conflict. For example, productivity shocks[1] induced by unexpected and exogenous weather events or changes in commodity prices have been found to increase the likelihood of a civil conflict onset (for a survey, see, for example, Blattman and Miguel (2010) and Burke et al. (2015)). Adverse productivity shocks have also been shown to increase the probability and intensity of other types of conflict behavior, such as stealing (Bignon et al. (2017)), or property and violent crimes (Papaioannou (2017), Miguel (2005)). The mechanism behind this regularity is, however, puzzling. As prominently formulated in Fearon (2008), while on the one hand, decreased ability to generate income from productive activity reduces the cost of engaging in conflict, on the other hand, also the value of everything that can be acquired in the conflict diminishes. As these two effects tend to offset each other, the observed systematic positive influence of adverse productivity and income shocks requires an explanation beyond the commonly cited mechanism relying on the opportunity cost of fighting.

I rely on the setting of the Hirshleifer-Skaperdas guns-and-butter game (Hirshleifer (1988), Skaperdas (1992)) as a model of conflict and extend the basic model by introducing loss preferences. I show in a theoretical model that an unexpected negative shock in productivity is expected to increase (and a positive one to decrease) conflict investments for loss-averse agents. The opposite holds for gain-seeking individuals, that is, those who show higher sensitivity to gains than to equal-sized losses.

[1]Here, as well as in the first chapter, I term under productivity shock an unexpected change in the ability to generate profit from productive (rather than appropriative activities. Although changes in income and productivity might refer to different concepts, in the model that I present the productivity of productive sector uniquely determines income.)

In a second step, I report the results of a conducted lab experiment ($N = 496$), with participants playing a repeated guns-and-butter conflict game under different productivity regimes. Particularly, in the two administered experimental treatments, an unexpected negative (respectively positive) productivity shock was introduced after several initial rounds were played at a given level of productivity. Participants in the control treatment continue playing according to the initial productivity level throughout the whole game.

The experimental results show that, in accordance with theoretical predictions, after a negative productivity shock, loss-averse individuals increase their conflict investments, and gain-seeking participants decrease them. However, while positive productivity shock leads to the predicted increase in conflict investments among gain-seeking participants, it fails to reduce conflict investments among loss-averse ones. Finally, independent of productivity level and in the absence of any productivity changes, a higher degree of loss aversion is associated with higher conflict investments.

In the second chapter, I study more subtle expressions of antagonism, namely prejudice expression and discrimination, which have been shown to both lead to and result from conflicts (DellaVigna et al. (2014), Shayo and Zussman (2011), Yanagizawa (2009)). In this part, I analyze immigration attitudes expressed by established immigrants and the sensitivity of these positions to groups' status positions. The work is motivated by a seemingly counterintuitive observation of the opposition to immigration to the host country, expressed by individuals who themselves immigrated some time ago. The support that Donald Trump's platform garnered among Latinx voters (in both 2016 and 2020), or the support that the Alternative für Deutschland received among immigrants from Ex-Soviet Union in Germany, exemplify the issue. I hypothesize that relative status deprivation, that is, the negative difference in status between own ethnic/national group and that of the native majority (or other, more favorably perceived minorities), has a negative impact on group's members' attitudes toward an even lower ranked status group (e.g., refugees). To investigate this idea, I use a survey experiment with a sample of participants with an immigration background residing in Germany (N=1,159) and experimentally vary the status of the participants'

in-group. Participants are presented a selected set of answers collected in a separate pre-study, where a sample of German participants with no immigration background was asked to evaluate the impact of a number of immigrant groups on the socio-economic and cultural life in Germany. The answers are selected such that the participant's in-group is evaluated positively in one and negatively in the other treatment while keeping constant the evaluations of other immigrant groups. The treatment variation, therefore, relies on manipulating the status position of participants in-group in the fixed ethnic hierarchy.

In line with the prediction, experimental results show that the support for refugees, captured by the willingness to forgo a part of experimental earnings in order to secure a donation to the United Nations High Commissioner for Refugees, was significantly lower in the treatment where participants received a negative evaluation of the own in-group. Furthermore, I show that exposure to prejudice leads to a change in perceived descriptive norm surrounding prejudice expression. More precisely, participants in the Negative treatment expect a significantly higher share of the majority population participants to have evaluated negatively (rather than positively) the impact of refugees from the Middle East. I provide evidence that the behavioral change (in willingness to donate) caused by treatment variation is channeled through altering this norm. These results provide support for the hypothesis that exposure to prejudice leads individuals to perceive prejudice expression as more frequent in the host society, reducing, in turn, their cost of expressing prejudice. Additionally, the role of indirect reciprocity as a possible moderator of observed treatment effects is considered. The results show that individuals with a higher preference for indirect reciprocity commit to donate more (compared to those with lower reciprocity preference) if they are allocated to the Positive treatment and less if they are in the Negative treatment.

Finally, the third chapter, based on joint work with Andrea Martinangeli, takes a step away from the conflict analysis and focuses more closely on the status and its consequences. We build upon the findings pointing at the role of status in the cultural evolution of human societies. The theory of social learning (Henrich and Gil-White (2001)) describes the evolution of status and prestige as a consequence of selection

favoring those individuals who were able to copy the most successful models in an efficient way. In turn, the most efficient in this social learning process were those individuals who, in order to enhance their learning environment, were ready to provide benefits and deference to successful models so as to achieve greater access and cooperation with them. The most successful models are therefore likely to have the largest following clienteles. Therefore, selection favors those followers who chose the model to follow on the basis of their current following clientele, and only with the time and additional information revelation refine their assessments of the relative model worth.

Relying on these ideas and the observation that status conferring features are sometimes distributed in a noisy manner (such as, e.g., through heritage), we hypothesize that monetary status, as one of the best recognizable status cues, might secure its owner preferred influence over others' behaviors, even in the situation where it does not signal superior task ability. In other words, we investigate whether there exists an influence premium of high monetary status and, if so, to which degree it depends on the visibility of an underlying knowledge or skill-based source of the status.

We study these questions in an online experiment where a representative sample of the German population (with respect to gender, age, income, and geographic area) is asked to decide the proportion of their experimental endowment that they want to invest in a lottery, which either triples or destroys the investment. Before making the investment decision, participants are offered advice left by an "advisor", a participant who took part in the same lottery at an earlier point of time. We assign to each advisor an endowment that can be either low or high, and that was allocated either randomly or according to advisor's result in a cognitive test, and present their advice to "advisees" together with this information.

Our findings support the hypothesis that high status, even when a product of chance, accords its possessor a strong influence premium over others' choices. Particularly, the results demonstrate that advice following is systematically higher if the advisor was highly endowed, in the case where the endowment was assigned randomly to the advisors. We further show that the result is driven by those advisees who achieved lower scores in the

cognitive test and had lower education levels. On the other hand, we detect no difference following between high- and low-endowed advisors if advisors "earned" the endowment by their performance in the cognitive test.

Chapter 1

Productivity Shocks and Conflict

1.1 Introduction

Do unexpected changes in productivity have an effect on the propensity to engage in conflict behavior or on the intensity of a conflict, and if so, what are the mechanisms behind this relationship? In the past decades, analysis of conflict and its relationship with income and productivity became a subject of study of many works in both theoretical and empirical economics. For example, a number of works studying civil wars, used weather conditions or international prices of exported goods as instruments for income to show that reduced ability to generate income in a productive sector led to an increased likelihood of suffering a civil conflict.[1] Adverse productivity shocks have also been shown to increase the probability and intensity of other types of conflict behavior, such as stealing (Bignon et al. (2017)), property crimes (Papaioannou (2017)), violent crime (Miguel (2005)), land invasions (Hidalgo et al. (2010)), communal conflicts (Bohlken and Sergenti (2010)), or coups (Kim (2016)). Furthermore, negative income shocks have also been found to increase the conflict intensity and its persistence (Bazzi and Blattman (2014), Chaudoin et al. (2017)). In a different setting, several works show that increased drug enforcement efforts on the part of the state, while effectively acting as a productivity shock from the perspective of drug trafficking groups, lead to an increased intensity of conflicts among them (Dell (2015), Werb et al. (2011)).

[1] The seminal paper by Miguel et al. (2004) introduced this approach, and was followed by a large set of works. For a survey of this literature see for example Blattman and Miguel (2010) and Burke et al. (2015)

The mechanism behind this regularity is however less clear. An often proposed explanation is based on the idea that a drop in income lowers the returns to the productive activity and thus reduces the agents' opportunity cost of engaging in predatory activities. However, as Fearon (2008) pointed out, this argumentation is at best incomplete, as with the decrease of income, the value of everything that can be captured in a conflict also decreases. Therefore, Fearon (2008) concludes that it is the second order effects of low income, namely the reduced ability of the state to tame the insurgencies or the lower ability to hide income from insurgents, that might help to explain the empirically supported relationship between low income and conflict.

This paper, starting from the behavioral perspective, provides additional insight into the relationship of income and conflict, using the setting of Hirshleifer-Skaperdas guns-and-butter game (Hirshleifer (1988), Skaperdas (1992)). In this type of conflict model, a player faces a trade-off between producing consumable goods and appropriating the production of other players (and defending their own production). I extend the basic model by introducing loss preferences and focus on analyzing the consequences of an income shock, which I model as an unexpected change in the marginal return of the productive activity. By incorporating loss preferences in the model of conflict, I allow for backward-looking behavior, as the agents reference their consumption to the previously held expectations. I show in a theoretical model that for loss-averse agents, an unexpected negative shock in productivity is expected to increase (and a positive one to decrease) conflict investments.

In a second step, I report the results of a conducted lab experiment ($N = 496$), with participants playing a repeated guns-and-butter conflict game under different productivity regimes. Particularly, in the two administered experimental treatments, an unexpected negative (respectively positive) productivity shock was introduced after several initial rounds were played at a given level of productivity. Participants in the control treatment continue playing according to the initial productivity level throughout the whole game. Additionally, participants' loss preferences were elicited.

As one of the most extensively tested non-standard preferences proposed by the prospect theory, loss aversion has proven helpful in explaining a broad spectrum of

economic behaviors. I study the implications of loss aversion for conflict behavior by allowing for expectation-based loss aversion a la Kőszegi and Rabin (2006) on the part of the players. Here, reference consumption itself is allowed to be stochastic, and thus agents form a reference lottery, which depends on the expectations that they have regarding the choices that they will face. Importantly, the reference lottery is based on the lagged expectations, rather than expectations contemporaneous with the time of consumption, which as Kőszegi and Rabin (2006) argue, is not suggesting that beliefs are slow to adjust, but rather that preferences do not instantaneously change when beliefs do. Thus someone expecting to play a certain lottery would experience a sensation of loss if this lottery was surprisingly substituted by a stochastically dominated lottery.

Applying this idea to the Hirshleifer-Skaperdas conflict model, I calculate the symmetric equilibrium conflict investments in the case where the agents expect the level of productivity according to which they are playing the conflict game, but also in the cases where the productivity shock creates a discrepancy between the expected conditions and those that are actually played. I show that in the guns-and-butter game, higher loss aversion is expected to lead to higher conflict investments for any given productivity level. Furthermore, an unexpected negative (positive) productivity shock is expected to increase (decrease) optimal conflict investments of loss-averse players. The opposite applies for the gain-seeking participants (that is, those attaching higher (absolute) utility to the gains relative to the same sized losses), for whom a negative income shock is predicted to lead to a decrease and a positive one to an increase in conflict investments.

Experimental results demonstrate that, in accordance with the theoretical predictions, an unexpected productivity drop increases conflict investments among loss-averse individuals and decreases them among gain-seeking ones, while loss-neutral participants show no change in conflict behavior. However, while a positive productivity shock does lead gain-seeking individuals to significantly increase their conflict investments, it fails to decrease conflict investments among loss-averse participants.

In response to the puzzle posed by the offsetting effects of adverse income shock on the likelihood of civil conflict onset, namely the decreasing opportunity costs conflict on the one side and decreasing value of the prize on the other, Fearon (2008) proposes that it is the second order effects of low income, namely the reduced ability of the state to tame the insurgencies or the lower ability to hide income from insurgents, that might help to explain the effect of income shocks. In addition to this explanation, several other works suggested theories of the influence of an income shock on conflict behavior while explicitly taking into account the opposing effects of income shock on the opportunity cost of conflict and the prize of conflict winner. For example, Chassang and Padro-i Miquel (2009) explain increased conflict propensity after a negative income shock as investing in conflict while the economy is temporarily low so as to secure control over a share of lootable resources, counting with the prospect of future economic recovery. Thus in their model, forward-looking behavior leads to the fact that, while the opportunity cost of fighting temporarily decreases, the prize allocated to the conflict winner decreases proportionately less, which in turn leads to higher incentives for conflict. Dal Bó and Dal Bó (2011) instead consider an economy consisting of two sectors - a capital-intensive and a labor-intensive one. They show in a theoretical model that a positive economic shock in the capital intensive sector can increase the value of control over the state without increasing the wages in the labor-intensive sector (and the other way around holds in labor intensive sector - negative shock decreases the wages without reducing the value of state control). Thus, positive income shocks in a capital intensive sector and negative ones in labor intensive sector can both lead to an increase in the probability of conflict.

This paper complements this literature by taking a different approach and focusing on the role of non-standard preferences, and particularly on reference dependence while leaving all other elements of the model as simple as possible. In both theoretical and experimental analysis, I consider a conflict game with symmetric players (thus with no differentiation between the state and insurgents or between economic sectors) and without future considerations. I thus abstract away from asymmetries in shock effects and from any second order effects of income shocks, such as the reduced state capacity, ability to hide income, or expectations of a rebounding economy. Therefore, rather than providing

an alternative model of a given conflict situation, the main focus of this work is to explore the role of loss aversion, as a potential additional and complementary channel, in an otherwise relatively simple setting of game of conflict.

Furthermore, this study also contributes to the experimental literature on conflict (for a recent survey, see Kimbrough et al. (2020)). Within this literature, where conflict is often modeled as a contest, the influence of loss preferences has already been examined in the context of contest games with an exogenous prize. Cornes and Hartley (2003) provide the first theoretical model that incorporates loss aversion in a Tullock contest (Tullock (1980)), and show that when allowing the agents to be loss averse around initial endowment, a higher level of loss aversion leads to lower equilibrium contest efforts. Experimental results of Kong (2008), Sheremeta (2015) and Shupp et al. (2013) confirm this theoretical result and find that more loss-averse individuals invest less in a Tullock contest. In a somewhat different approach, Chowdhury et al. (2018) consider the role of loss aversion in a Tullock contest, conditional on whether the players are initially holding property rights over the contest prize, which they stand to lose in case of losing the contest. They show, both theoretically and in an experiment, that when property rights are initially given to the players, higher loss aversion entails higher contest investments. However, other than employing a model of loss preferences different than the one used in this paper, these works also do not consider income shocks and their impact on conflict behavior. On the other hand, income shocks have seldom been considered within experimental literature on conflict. One exception is Baik et al. (2020), who, modelling conflict as Tullock contest, study the effect of varying the size of conflict budget (while holding equilibrium contest investments constant) on investments in conflict.[2] They find a non-monotonic relationship, with conflict investments increasing in the budget up to some point but decreasing thereafter. In a related study, Price and Sheremeta (2011) find higher overbidding in the contest, when contestants are given the total budget for all rounds of the game, than in the case where

[2]As cited in the Baik et al. (2020), a few other works - including Morgan et al. (2012), Sheremeta (2013), Sheremeta (2011), Sheremeta (2010) - while primarily focusing on other aspects, report a positive effect of contest budget on investments in the contest.

the budget is equally split between the rounds and provided to players piece-wise. These studies, however, do not consider the loss preferences of players.

Finally, this works also contributes to the narrower subset of experimental conflict literature, namely those works considering the guns-and-butter game of conflict. Previous experimental studies of conflict in the the guns-and-butter game setting predominantly explored the behavioral consequences of asymmetry in players' strength (Kimbrough et al. (2014)), including "paradox of power" (Durham et al. (1998)), the effects of post-conflict behavior and repeated interaction (Lacomba et al. (2014)), as well as the role of consideration of future (McBride and Skaperdas (2014), Tingley (2011)). To the best of my knowledge, this is the first work that studies the implications of both loss aversion and income shocks in the guns-and-butter model of conflict.

The rest of the paper is structured in the following way. Section 2 introduces the model of the guns-and-butter conflict game and provides a theoretical analysis of the impact of loss aversion in this game. Section 3 describes the conducted experiment. Section 4 presents the analysis of experimental results. Section 5 concludes.

1.2 Guns-and-butter game

The conflict game analyzed here is a simple version of the Hirshleifer-Skaperdas guns-and-butter conflict game (Hirshleifer (1988), Skaperdas (1992)). Two rational, risk and loss neutral players, denoted by i and j, are endowed each with resources in value of $E > 0$. They make simultaneous decisions regarding how to split the endowment between the two investment possibilities that we will term as "conflict" (G) and "production" ($E-G$). Thereby each unit of endowment invested in production creates $\beta > 0$ units of consumable resource (i.e., the production technology is linear with marginal productivity equal to β). If any player invests a positive amount in conflict both players compete in a winner-take-all lottery contest, whereby, denoting investments in conflict of players i and j by G_i and G_j respectively, the player i's probability of winning equals the share of his conflict investment in the total conflict investments $p_i(G_i, G_j) = G_i/(G_i+G_j)$, and the prize of the contest equals the value of total production of both players $P(G_i, G_j) = \beta(2E-G_i-G_j)$.

If both players invest zero in conflict, each of them keeps their own production of βE. Thus, the expected utility of player i can be written as:

$$E(V(G_i, G_j)) = G_i/(G_i + G_j)\beta(2E - G_i - G_j) \tag{1.1}$$

and the unique symmetric Nash equilibrium investments in conflict by both players equal:

$$G_i^* = G_j^* = G_{NE}^* = E/2 \tag{1.2}$$

Denote this optimal investment of a risk and loss neutral player who maximizes own expected payoff as G_{NE}^*. It is immediately clear that the peace solution, $G_i^* = G_j^* = 0$ cannot be the equilibrium, as both players would have an incentive to deviate by investing a strictly positive and arbitrarily small ε in conflict and thus win the conflict for sure with an arbitrarily small cost. What also follows directly from the G_{NE}^* formula is that the optimal investment does not depend on the productivity level β and thus any change in productivity level should leave the optimal investments in conflict unchanged. This is also intuitive since any change in productivity would lead to two offsetting effects – an increase (decrease) of productivity leads to an increase (decrease) of the opportunity cost of investing in conflict, but at the same time to an increase (decrease) of value of the contest prize.[3]

The optimal investment in conflict as derived above is based on the assumption that all players are perfectly rational, risk and loss neutral maximizers (and that this is common knowledge). However, considerable experimental evidence from (among others) distinct

[3]As emphasized there, this result also holds for a number of different specifications of the model. For example, if the initial income would not be divided equally among the players, so that $E_i \neq E_j$, optimal investments in conflict for both players would equal $(E_i + E_j)/4$ and would therefore again remain independent of the productivity level. Alternatively, if the productivity would not be the same for both players, so that $\beta_i \neq \beta_j$, then in the general case also the optimal investments in conflict would be different for the two players. However, applying the same proportional change to both productivity levels (that is, multiplying both β_i and β_j by the same factor $\gamma > 0$) leaves the optimal investments in conflict unchanged. Finally, allowing for the risk aversion also does not necessarily change this result since the optimal conflict investment stays independent of the productivity level for any CRRA utility function. Therefore, a specific form of risk aversion (particularly, increasing relative risk aversion) would have to be assumed so as to explain the negative correlation between productivity (β) and conflict ($G_i^* + G_j^*$).

but related contest games suggests that these assumptions might not be satisfied and that different behavioral biases might affect behavior in guns-and-butter game of conflict. In the following, I allow for the backward-looking behavior by analyzing the effects of loss aversion, where the reference consumption is determined by the conditions experienced in the recent past.

1.2.1 Loss aversion

Consider again the same guns-and-butter game, but now allowing for the loss aversion. Specifically, assume that the utility function common to both players i and j depends on both consumption level c (where c is the value of resources available for consumption after resolution of uncertainty) and a reference point R, and can be represented as:[4]

$$V(c, R) = c + \eta v(c - R)$$

where

$$v(c-R) = \begin{cases} c - R, & \text{if} \quad c - R \geq 0 \\ -\lambda \left| c - R \right|, & \text{if} \quad c - R < 0 \end{cases}$$

Here, λ is the parameter of loss aversion, and η is a weight attached to gain-loss utility (relative to the consumption utility). Thus, $\lambda > 1$ would imply loss aversion, that is, attaching higher (absolute) utility to the losses than to the gains of the same size. Equivalently, $\lambda < 1$ would imply gain seeking, i.e., attaching higher (absolute) utility to the gains relative to the same sized losses, and $\lambda = 1$ would indicate loss neutrality.

In a difference to the fixed-prize contest games, where players keep for sure the part of the endowment that they haven't invested in contest, in the case of a guns-and-butter game, there exists no safe haven. The part of the endowment that is not invested in guns is invested in producing butter, which in turn forms a total prize that is allocated to the winner of the contest. In other words, no investment decision can guarantee any

[4] As above, I am assuming here a linear consumption utility.

fixed value of consumption (including zero), which makes the concept of a fixed scalar reference point difficult to apply in this setting. A theory developed by Kőszegi and Rabin (2006) and Kőszegi and Rabin (2007) provides a modeling framework that allows the referent to be stochastic as well, so that the gains and losses are evaluated against a reference lottery rather than a reference point. Particularly, when playing a given lottery $L = \{c_i : p(l_i)\}_{i=1}^{N}$, any outcome c_i is evaluated by comparing it to all possible outcomes in the support of the reference lottery $L_R = \{r_j : p(r_j)\}_{j=1}^{M}$, and is further weighted by the respective probabilities of the both outcomes that are being compared (that is $p(c_i)$ and $p(r_i)$). Therefore, the expected utility can be represented as:

$$E[V(c,R)] = \sum_{i=1}^{N}\sum_{j=1}^{M} V(c_i, r_j)p(c_i)p(r_j) = \sum_{i=1}^{N}\sum_{j=1}^{M}(c + \eta v(c_i - r_j))p(c_i)p(r_j) \tag{1.3}$$

The reference lottery L_R in the theory of Kőszegi and Rabin is based on the rational lagged expectations of the distribution of outcomes that an individual will be facing. More precisely, the concept of choice-acclimating personal equilibrium (CPE), as defined in Kőszegi and Rabin (2007), assumes that individuals can correctly predict the choices that they will be facing, the set of possible outcomes, and how the distribution of these outcomes depends on their actions. In this case, a player who expects to play a lottery L should also hold the same lottery L as a reference lottery and evaluate any given outcome by comparing it to all possible outcomes in the support of L. However, in the case where a player faces an ex-ante low probability situation, the reference lottery can be regarded as set by the previously held expectations and thus fixed, basically irrespective of the newly relevant choice set. As Kőszegi and Rabin (2007) argue, this is not suggesting that beliefs are slow to adjust, but rather that preferences do not instantaneously change when beliefs do. In other words, if a player rationally expected to play lottery L, so that the lottery L was also a reference lottery of that player, and then was surprisingly faced with playing a different lottery L', then at least for some time, their reference lottery would remain fixed at L.

In the context of the guns-and-butter game, denote by L_h the game as presented above, where the productivity factor β is set at some value β_h, and equivalently denote by L_l the same game, but where the productivity factor β is set to equal β_l, where $\beta_h > \beta_l$. For given conflict investments, the probability of winning the conflict remains unchanged in both games, but the prize allocated to the winner is higher in L_h than in L_l, whereas the loser receives zero in any case. Denote by $E[V(G_i, G_j, L, L')]$ the expected utility of the player who is playing the lottery L whereas their reference lottery is lottery L'. In the following subsections, I consider the case where, due to the unexpected decrease (increase) in the marginal productivity of butter, the reference lottery remains fixed at L_h (L_l), but the actually played lottery changes to L_l (L_h), as well the one where the reference lottery coincides with the lottery that is actually played, and determine the optimal conflict investments.

Negative productivity shock: Reference lottery stochastically dominates the played lottery

Consider the case where both players were introduced to game L_h, which they expected to continue playing before the unexpected productivity change from β_h to β_l is introduced (where $\beta_h > \beta_l$), so that as of that moment, they switch to playing the game L_l. If agents' preferences adjust to the new conditions with some delay relative to their beliefs, the previous productivity level (β_l) continues to enter agents' reference lottery for some time after the productivity shock. Agent i's expected utility can be written as:

$$E[V(G_i, G_j, L_l, L_h)] = \frac{G_i}{G_i + G_j}\beta_l\left(2E - G_i - G_j\right) + \eta(-\lambda\frac{G_i^2}{\left(G_i + G_j\right)^2}(\beta_h - \beta_l)\left(2E - G_i - G_j\right) + \frac{G_iG_j}{\left(G_i + G_j\right)^2}\beta_l\left(2E - G_i - G_j\right) - \lambda\frac{G_iG_j}{\left(G_i + G_j\right)^2}\beta_h\left(2E - G_i - G_j\right))$$

The first term in the utility expression represents the usual consumption utility of the winner's prize $\beta_l(2E - G_i - G_j)$ weighted by the probability of winning. The second term

(in parenthesis) is the expected gain-loss utility of lottery L_l relative to the reference lottery L_h. As with the decrease of the productivity also the winner's prize decreases, winning in the game L_l, while holding L_h as a reference lottery, would lead to both loss and gain sensations. Particularly, the first summand represents expected loss, which results from comparing the outcome of winning in L_l with that of winning in L_h, $((\beta_h - \beta_l)(2E - G_i - G_j))$, amplified by the loss aversion coefficient λ and weighted by the probabilities of the two outcomes. In a similar way, the second summand represents the expected gain resulting from comparing the utility of winning in the lottery played with that of losing in the reference lottery. The last summand captures the expected loss, resulting from comparing utilities of losing in the L_l and winning in L_h. Finally, η is the weight assigned to the gain-loss utility (relative to the consumption utility, which is weighed by 1). Throughout the paper, I assume that the gain-loss utility weight and the coefficient of loss aversion are strictly positive ($\eta > 0$, $\lambda > 0$).

Note that for any interior solution (G_i, G_j) the gain-loss utility is negative for $\lambda > 1$. For sufficiently high values of η and λ then, the impact of gain-loss utility would become very high relative to the consumption utility, and the utility-loss due to the gain-loss utility would overweight any utility-gain from consumption utility, so that the total utility would be negative for any symmetric interior solution (and deviating to playing $G_i = 0$ so as to secure zero utility would become profitable). I restrict the values of parameters so as to focus only on interior symmetric solutions, which here implies assuming:

$$\frac{\beta_h}{\beta_l} < \min\{\frac{\eta + 3\eta\lambda + 4}{4\eta\lambda}, \frac{2\eta + 2\eta\lambda + 4}{4\eta\lambda}\} \tag{1.4}$$

where both terms on the right side of the inequality decrease in η and λ. The unique interior symmetric choice-acclimating personal Nash equilibrium, denoted by $G^*_{l,h}$, in this case equals:

$$G^*_i = G^*_j = G^*_{l,h} = 2E\frac{\beta_l(1 + \eta\lambda) - \eta\lambda\beta_h}{\beta_l(\eta + 3\eta\lambda + 4) - 4\eta\lambda\beta_h} \tag{1.5}$$

Positive productivity shock: Reference lottery is stochastically dominated by the played lottery

Consider now the case where both players, after playing the game L_l for some time, and thus having adopted it as a reference game, unexpectedly switch to playing L_h. The expected utility of player i in this case equals:

$$E[V(G_i, G_j, L_h, L_l)] = \frac{G_i}{G_i + G_j}\beta_h\left(2E - G_i - G_j\right) + \eta\left(\frac{G_i^2}{\left(G_i + G_j\right)^2}\left(\beta_h - \beta_l\right)\left(2E - G_i - G_j\right) + \frac{G_i G_j}{\left(G_i + G_j\right)^2}\beta_h\left(2E - G_i - G_j\right) - \lambda\frac{G_i G_j}{\left(G_i + G_j\right)^2}\beta_l\left(2E - G_i - G_j\right)\right)$$

Switching to playing the lottery that stochastically dominates the reference lottery adds an additional gain component to the expected gain-loss utility, because the prize that the conflict winner receives is higher than would be the case in the reference lottery. Consequently, the gain-loss utility is now not necessarily negative in the general case, but as above, for very high values of η and λ relative to the size of productivity shock ($\frac{\beta_h}{\beta_l}$), the perspective of possible loss would weigh so high, that the total utility would become negative for any interior symmetric solution. As above, I restrict the values of the parameters so as to focus only on interior symmetric solutions, which implies:

$$\frac{\beta_h}{\beta_l} > \max\{\frac{3\eta + \eta\lambda}{4\eta + 4}, \frac{2\eta + 2\eta\lambda}{4\eta + 4}\} \tag{1.6}$$

where both terms on the right side of the inequality increase in η and λ. The unique interior symmetric choice-acclimating personal Nash equilibrium in this case, denoted by $G^*_{l,h}$, equals:

$$G_i^* = G_j^* = G_{h,l}^* = 2E\frac{\beta_h(1 + \eta) - \eta\beta_l}{\beta_h(4\eta + 4) - (3\eta + \eta\lambda)\beta_l} \tag{1.7}$$

Reference lottery coincides with the played lottery

Finally, the expected utility of player i, playing with player j the game L_l, in the absence of any productivity shock, that is when both players expect to play the same game (and thus have it as their reference game), can be written as:

$$E[V(G_i, G_j, L_l, L_l)] = \frac{G_i}{G_i + G_j}\beta_l (2E - G_i - G_j) + \eta \left(\frac{G_i G_j}{(G_i + G_j)^2}\beta_l (2E - G_i - G_j) - \lambda \frac{G_i G_j}{(G_i + G_j)^2}\beta_l (2E - G_i - G_j) \right)$$

As above, I restrict the values of parameters so as to focus only on interior symmetric solutions, which here implies assuming:

$$\eta(\lambda - 1) < 2. \tag{1.8}$$

The unique interior symmetric choice-acclimating personal Nash equilibrium in this case, denoted by $G^*_{l,l}$, equals:

$$G^*_i = G^*_j = G^*_{l,l} = 2E\frac{1}{4 - \eta(\lambda - 1)}. \tag{1.9}$$

Thus, if both players' reference lotteries coincide with the lottery that they indeed are playing, the optimal conflict investment is independent of the productivity level β (implying $G^*_{l,l} = G^*_{h,h}$)[5] and is increasing in loss aversion.

1.2.2 Comparative statics and hypothesis

Based on the results derived above, we can form the predictions regarding conflict investments in the aftermath of an unexpected productivity change and conditional on the sign and intensity of loss preferences. Particularly, comparing optimal conflict

[5]In the further text, I will always denote optimal investment in the case where reference lottery coincides with the lottery that is played by $G^*_{l,l}$ to avoid excessive use of notation.

investment after a negative productivity shock $G^*_{l,h}$ (equation 1.5) and the optimal conflict investment in absence of the shock $G^*_{l,l}$ (equation 1.9), it can be shown that:

$$\begin{aligned} G^*_{l,h} > G^*_{l,l}, \quad \text{if} \quad \lambda > 1 \\ G^*_{l,h} < G^*_{l,l}, \quad \text{if} \quad \lambda < 1. \end{aligned} \tag{1.10}$$

Furthermore, the difference between the optimal conflict investment after a negative productivity shock and prior to it $(G^*_{l,h} - G^*_{l,l})$ can be shown to increase in λ. These results form a base for the first hypothesis:

Hypothesis 1: Conflict investments in a guns-and-butter game, after an unexpected negative productivity shock, are expected to increase for loss-averse players ($\lambda > 1$) and to decrease for gain-seeking players ($\lambda < 1$), relative to the level prior to the shock. Furthermore, the size of the increase (decrease) is expected to be larger the more loss-averse (gain-seeking) the player is.

Similarly, comparing the optimal conflict investment after a positive productivity shock $G^*_{h,l}$ (equation 1.5) and the optimal conflict investment in absence of the shock $G^*_{l,l}$, it can be shown that:

$$\begin{aligned} G^*_{h,l} < G^*_{l,l}, \quad \text{if} \quad \lambda > 1 \\ G^*_{h,l} > G^*_{l,l}, \quad \text{if} \quad \lambda < 1. \end{aligned} \tag{1.11}$$

Moreover, the difference between the conflict investment prior to and after the productivity shock can be shown to increase in the level of loss aversion λ. The combination of these results leads to the second hypothesis:

Hypothesis 2: Conflict investments in a guns-and-butter game, after an unexpected positive productivity shock, are expected to decrease for loss-averse players ($\lambda > 1$) and to increase for gain-seeking players ($\lambda < 1$), relative to the level prior to the shock. Furthermore, the size of the decrease (increase) is expected to be larger the more loss-averse (gain-seeking) the player is.

Finally, from equation 1.9 it follows directly that the optimal conflict investments in the absence of any productivity shock, that is where the reference lottery of a player coincides with the lottery played, is independent of the productivity level β (implying $G^*_{l,l} = G^*_{h,h}$) and is increasing in loss aversion. This leads to the third hypothesis:

Hypothesis 3: Denoting optimal conflict investments in a guns-and-butter game, where the reference lottery of a player coincides with the lottery that they are actually playing by $G^*_{l,l}$, and assuming that the weight assigned to the gain-loss utility is positive $\eta > 0$:

a) $G^*_{l,l}$ is independent of the productivity level β

b) $G^*_{l,l}$ is increasing in the level of loss aversion λ, for any $\lambda > 0$.

One remark is in order here. Namely, as the action set of players here consists of the amount they invest in guns, there is no possibility for a player to invest in guns but refrain from attacking the other side. Thus, in instances where there exists a conceptual difference between the outbreak of a conflict and its consecutive intensity, this game is more accurately modeling the latter than the prior. However, as mentioned above, the influence of income shocks has been found to be significant also in the case of intensity of the already existing conflicts. Furthermore, some sorts of extractive and conflict behavior do not allow the separation of the conflict investments and their implementation. For example, while it is obvious that arms can be held for deterrent purposes, investing in property crime or robbery mostly coincides with its realization.

1.3 Experimental Design and Procedures

In order to examine the behavior in the described conflict game, a lab experiment was conducted, involving a total of 496 subjects, divided into 27 sessions. All sessions were conducted between October 2019 and January 2020 at the econlab laboratory in Munich, Germany. The program was coded in z-Tree (Fischbacher (2007)), and the subjects were invited through ORSEE (Greiner (2015)). Although there were no restrictions on age

or occupation, most of the participants were local students.[6] The subjects were told that an experiment would consist of multiple parts but were not aware of the content of any part before its beginning. The absolute strangers matching mechanism was applied in the parts of the experiment where the subjects interacted in pairs. In the first part of the experiment (guns-and-butter game) these pairs were drawn from the sets of 8 participants, and the randomization is done on the level of those groups. Each session lasted approximately 90 minutes.

In the first part of the experiment, participants played seven rounds of the guns-and-butter game as described above. Prior to playing the game, participants received the instructions[7] (that were provided on paper and were also read out loud by an assistant). The instructions explained that seven rounds would be played and that in each round, the participants would receive an endowment of 100 tokens that they have to allocate between investments in project P1 and project P2. Each token invested in project P1 generated β many points, and each point was equivalent to 0.01 EUR. Investments in project P2 were explained to determine the probability of winning the total points produced by both players, whereby the probability of winning of any player equals the share of that player's P2 investment in the total P2 investment. The winner was determined via a wheel of fortune. The instruction also stated that in the case that both players invested zero tokens in project P2, both players would keep all the points that they generated in project P1, and equivalently that if both players invested zero tokens in project P1, the winner would still be determined, but the prize would equal zero points. The participants were informed that one of the seven rounds would be randomly selected to be payoff relevant at the end of the experiment. The instructions also stated explicitly that in each round, each participant will be matched with another randomly chosen participant and that matching with the same participant in more than one round is not possible.

After reading the instructions, participants were shown two examples, with fictional players and randomly generated investments. The same examples were shown to all

[6] Roughly 56% of all participants were female. Slightly more than 93% of all participants were students, thereof around 37% studied math, engineering or natural sciences, 20% studied economics or business administration, and the rest was distributed among other disciplines.

[7] Instructions for all parts of the experiment are available in the online appendix.

participants within one session, and the subjects were informed that the examples were fictional and randomly determined. Thereafter, all participants were asked four comprehension questions that they had to answer correctly so as to proceed to play the first round of the game.

Each participant was allocated to one of the three treatments:

- **Control Treatment** (N=168) - In this treatment the productivity β was set to equal 6, that is, each token invested in project P1 generated 6 points.
- **Shock Down Treatment** (N=160) - In this treatment, the instructions (including examples and comprehension questions) were presented always using the productivity factor $\beta = 10$ and the first four rounds of the game were also played with this productivity level. After the fourth round participants were informed that, as of the next round, the productivity will be set at $\beta = 6$, and that it would stay on that level for the rest of the game.
- **Shock Up Treatment** (N=168) - In this treatment, the instructions (including examples and comprehension questions) were displayed and the first four rounds were played according to the productivity $\beta = 6$. After the fourth round, participants were informed that, as of the next round, the productivity will be set at $\beta = 10$, and that it would stay on that level for the rest of the game.

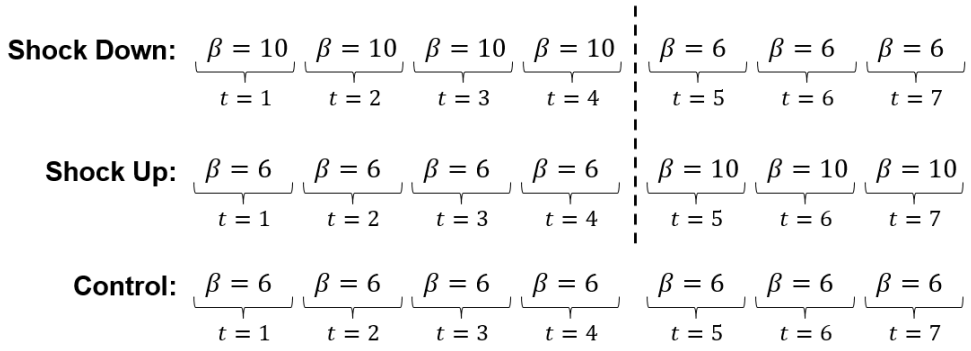


Figure 1.1. Treatments in the guns-and-butter game: productivity levels across time periods

Figure 1.1 provides a schematic representation of the played guns-and-butter game and the three treatments. The assignment to one of the three treatments was random and

made on the level of a session and the order of session-treatments was also randomized. Additionally, in each round of the game, each participant was asked to guess the P2 investment of the other participant, and an answer that was +/- 5 tokens away from the actual investment of the other player was rewarded by additional 100 points.

After each round of investments, the investments of both players and the wheel of fortune that determined the winner of the contest were displayed along with the calculated earnings of the participant from that round, including both earnings from the guns-and-butter game and those from guessing the opponent's P2 investment. As explained above, at the end of the experiment, one randomly selected round of the guns-and-butter game was chosen as payoff-relevant for this part of the experiment, and participants' earnings from this round (both from the guns-and-butter game itself and those from the guessing opponent's P2 investment) were added to their experimental earnings.

In principle, only two rounds in each treatment would have been sufficient to observe the potential change in behavior due to the productivity shock. However, in his experimental work, Heffetz (2018) suggests that reference points reflect sunk-in rather than merely time-wise lagged expectations. In a nutshell, this would imply that the expectations need to be somehow internalized, that is, the participants need to get used to them, so as to serve as a basis for forming a reference point. Thus, the design involving seven rounds of the guns-and-butter game was chosen as a trade-off between the attempt to allow for the sufficient number of rounds to facilitate the sink-in effect[8], and on the other side the need to account for potential clustering within a group from which the absolute stranger pairs are drawn.

Treatments differed only in the guns-and-butter game, and in the rest of the experiment participants in all three treatments completed the same tasks.

After completing the guns-and-butter game, a measure of joy of winning, that is, the utility that agents derive from winning per se, independent of the associated payoff, was elicited in accordance with the procedure developed by Sheremeta (2010), where the participants had the possibility to invest any non-negative number of tokens (from the

[8]To further strengthen the sink-in of the productivity level prior to the shock, all instructions and examples were using the initial level of productivity.

Table 1.1. Elicitation of loss preferences

	Option A	Option B
1	0 points for sure	50% chance to win 250 points; 50% chance to lose 25 points
2	0 points for sure	50% chance to win 250 points; 50% chance to lose 50 points
3	0 points for sure	50% chance to win 250 points; 50% chance to lose 75 points
4	0 points for sure	50% chance to win 250 points; 50% chance to lose 100 points
5	0 points for sure	50% chance to win 250 points; 50% chance to lose 125 points
6	0 points for sure	50% chance to win 250 points; 50% chance to lose 150 points
7	0 points for sure	50% chance to win 250 points; 50% chance to lose 175 points
8	0 points for sure	50% chance to win 250 points; 50% chance to lose 200 points
9	0 points for sure	50% chance to win 250 points; 50% chance to lose 225 points
10	0 points for sure	50% chance to win 250 points; 50% chance to lose 250 points
11	0 points for sure	50% chance to win 250 points; 50% chance to lose 275 points
12	0 points for sure	50% chance to win 250 points; 50% chance to lose 300 points
13	0 points for sure	50% chance to win 250 points; 50% chance to lose 325 points
14	0 points for sure	50% chance to win 250 points; 50% chance to lose 350 points
15	0 points for sure	50% chance to win 250 points; 50% chance to lose 375 points
16	0 points for sure	50% chance to win 250 points; 50% chance to lose 400 points
17	0 points for sure	50% chance to win 250 points; 50% chance to lose 425 points
18	0 points for sure	50% chance to win 250 points; 50% chance to lose 450 points
19	0 points for sure	50% chance to win 250 points; 50% chance to lose 475 points
20	0 points for sure	50% chance to win 250 points; 50% chance to lose 500 points

Note: The subjects choose between the safe payoff of zero points (option A) and a lottery entailing a chance to win or loose a given number of points (option B) in each of the 20 questions. The payoff from this part of the experiment is based on one, randomly selected question.

100 token endowment that was allocated for this task) in a zero-prize (two-player) Tullock contest.

In the following step of the experiment, participants had to provide answers to five cognitive reflection questions. The cognitive reflection test (CRT) was established by Frederick (2005) as a measure of the ability to override a seemingly intuitive but wrong response and to engage in further reflection in order to arrive at a correct response. In the context of this experiment, the CRT score was intended as a proxy measure for the ability of participants to see through the structure and understand the mechanics of the guns-and-butter game. The first three of the five questions that were presented were the (minimally modified) questions provided by Frederick (2005). In order to reduce the influence of previous experience with the questions, another two questions adopted from the set of questions provided by Toplak et al. (2014) were added, resulting in a total of five questions which are presented in the appendix. The participants were informed that one of the five questions would be randomly selected at the end of the experiment, and if

the participant provided the correct answer to that question, 150 points would be added to their earnings.

In the next part of the experiment, the subjects' preferences toward losses were elicited from a set of 20 questions as shown in Table 1.1. In each of the questions, the subjects were asked to state whether they preferred the safe option (receiving 0 points for sure) or the risky option (50% chance of losing a certain amount between 25 points and 500 points and 50% chance of winning 250 points). The subjects were further informed that at the end of the experiment, one of the questions will be randomly selected and, depending on the provided choice, either a lottery outcome or a safe option will determine the earnings from this part of the experiment. The ordering of the questions is such that more loss-averse subjects would choose more safe options than less loss-averse subjects. Particularly, noting that all lotteries in Table 1 have the form $L_n = \{250 : 0.5; -25n : 0.5\}$, where n is the number of the question in the table ($n \in \{1, 2, ..., 20\}$), a participant will decide for the lottery, in each question n where the expected utility of lottery L_n exceeds zero. The content of this part of the experiment was unknown to the participants before its start, and thus there are no reasons to believe that the participants were expecting to face the given lotteries in advance. Additionally, here, in contrast to the case of the guns-and-butter game, participants can secure a sure payoff of zero by completely refraining from playing the lotteries. Considering this, I argue that in this case, a sure zero payoff can be considered as the fixed reference point in the loss-aversion elicitation task.

Therefore, the expected utility of lottery L_n equals:

$$E[V(L_n, 0)] = \frac{1}{2}250 - \frac{1}{2}25n + \eta(\frac{1}{2}250 - \frac{1}{2}\lambda 25n) \tag{1.12}$$

A participant will choose a lottery in question n as long as[9]:

$$10(1 + \eta) \geq n(1 + \eta\lambda). \tag{1.13}$$

[9]I assume that when indifferent between a lottery and a sure payoff a participant chooses the lottery.

I denote the measure of loss aversion by LA and set it to equal the number of questions where a participant chose safe zero payoff diminished by 10. From equation 1.13 it follows that any loss averse participant ($\lambda > 1$) will switch to choosing a sure zero payoff at some question n, where $n < 10$ ($LA > 0$), whereas the gain-seeking participants will decide for lotteries up to some question n, where $n > 10$ ($LA < 0$). Finally, the loss-neutral participants are expected to choose the lottery in exactly 10 questions ($LA = 0$).

As usual, some subjects provided inconsistent answers to the lottery selection task, as they switched from lottery to safe choices more than once. Particularly, in the three treatment groups, Shock Down, Shock Up and Control, there are 9 (out of total 168), 16 (out of total 160) and 8 (out of total 168) subjects respectively, whose answers were inconsistent and whose data was consequently eliminated (thus leaving a total of 463 subjects).

The distribution of the loss aversion measure for all participants who provided consistent responses is shown in Figure 1.2. The distributions of LA in the three treatment groups did not differ significantly (Kruskal-Wallis test: $Chi\ square = 0.366$, $p = 0.833$; mean values of LA in Shock Down, Shock Up and Control group equal 10.82, 11.45 and 10.98 respectively) indicating successful randomization with respect to this parameter. Additionally, the number of rounds of the guns-and-butter game in which a participant won the conflict also had no influence on the behavior in the subsequent loss aversion test.

Finally, the participants also completed a questionnaire with standard demographic questions regarding their age, field of study, etc.

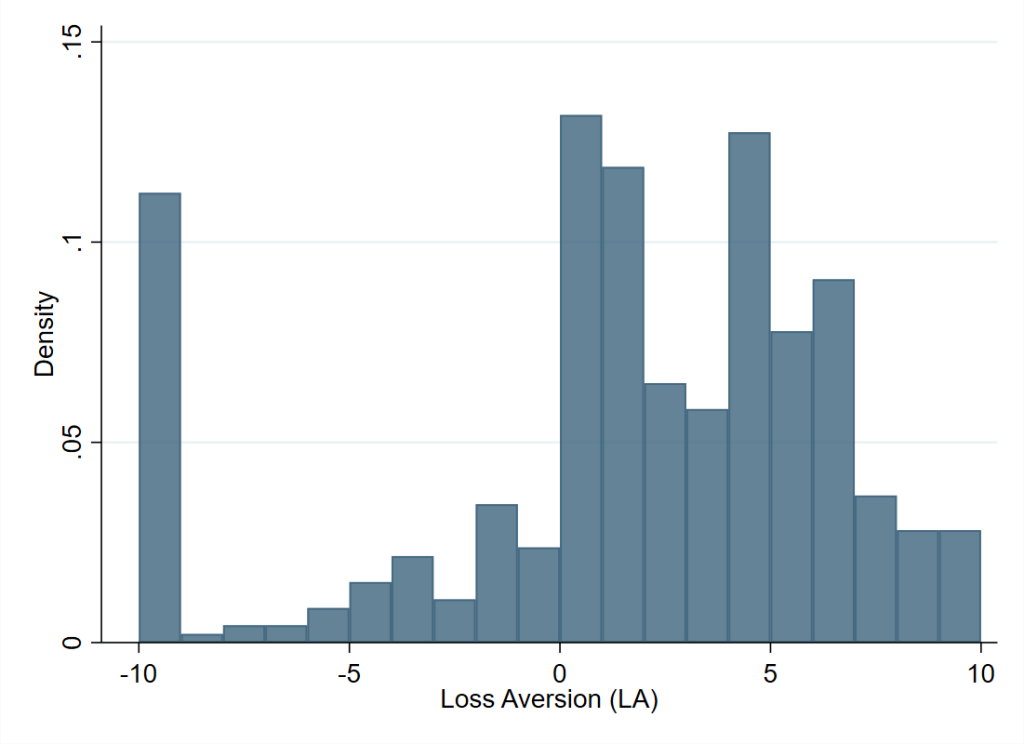


Figure 1.2. Distribution of measure of loss aversion (LA) across all participants from all three treatments (N=463). LA equals the number of safe choices selected from a set of 20 lotteries (as in Table 1).

1.4 Experimental Results

1.4.1 Average conflict investments

Figure 1.3 provides an overview of the average conflict behavior. The three panels of the figure depict average conflict investments for each round of the game and in each treatment. The average investment in conflict over all periods of the game and all treatments is 63.01 token, which is substantially higher than the theoretical equilibrium prediction if it is commonly known that all players are risk and loss neutral and maximize only their monetary earnings ($G_{NE} = 50$ token). This difference is significant (Wilcoxon Signed Rank (WSR) test, $p < 0.001$) and also holds if each of the treatments is considered separately.

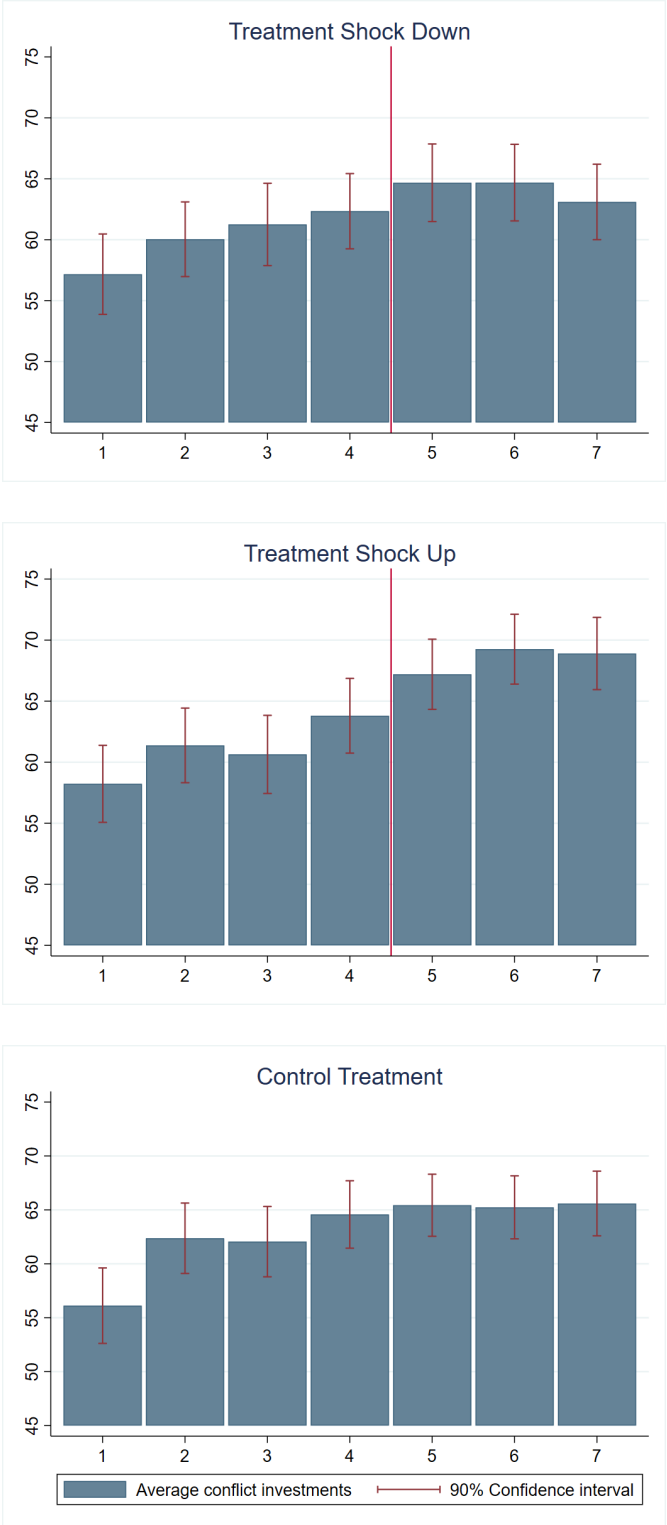


Figure 1.3. Average conflict investments in each of the seven rounds of the guns-and-butter game. The three panels depict the average conflict investments, separately for loss-averse participants ($LA > 0$) and for the gain-seeking ones ($LA < 0$), in each treatment.

1.4.2 Loss aversion and productivity shocks

In order to be able to account for the individual preferences of subjects, as well as for the structure of the game, in what follows, I use the GLS regressions with individual investments in conflict as the dependent variable. All of the regressions include subject random effects and the standard errors are clustered on the level of the groups (of 8 participants) from which the "absolute stranger" pairs were drawn.

Experimental results of Lacomba et al. (2014) demonstrate the impact of lagged behavior of participants on the dynamics of a repeated guns-and-butter game. Particularly, they show that lagging behind the opponent's conflict investments, as well as having lost the conflict in the previous round of the game, have a strong positive impact on participants' current conflict investments. Therefore, in order to be able to cleanly detect the impact of the shocks, in the regressions below, I include (unless where otherwise indicated) the set of variables capturing own conflict investment in the previous round of the game, the difference between the own and opponent's conflict investment in the previous round and a dummy variable indicating whether the participant won in the previous round's conflict. I collectively term this set of variables as "*Lagged behavior*" and postpone the discussion of their influence for the section "Escalation in Conflict Investments".[10]

Hypothesis 1 and Hypothesis 2 lay out predictions for the effects of a positive and negative productivity shock, respectively, conditional on the sign of loss preferences of a player (that is, whether they are loss-averse, gain-seeking, or loss-neutral) and on the intensity of those preferences. Particularly, fixing the productivity in the reference lottery at the level that applied prior to the shock, after an unexpected productivity decrease (treatment Shock Down), we expect to see an increase in conflict investments for loss-averse participants and a decrease in conflict investment for gain-seeking participants. Vice-versa applies after an unexpected productivity increase (treatment Shock Up), where we expect to see a decrease in conflict investment for loss-averse participants and an

[10]Whenever Lagged behavior is controlled for, the data of the first round of the game is omitted. This causes a loss of power, but at the same time ensures that the results are not driven by the variation in own or opponent's behavior from the previous round.

increase in conflict investment for the gain-seeking ones. In both cases, we expect no effect of shock for the loss neutral participants.

In order to perform a first test of the predicted results, two dummy variables, I_+ and I_- are created, which pool together all the participants for whom a positive (respectively negative) shock effect is predicted. More precisely, I_+ takes value one if a participant is loss-averse and is in the Shock Down treatment, or if a participant is gain-seeking and in the Shock Up treatment and zero otherwise. Equivalently, I_- equals one if a participant is gain-seeking and is in the Shock Down treatment, or if a participant is loss-averse and is in the Shock Up treatment and zero otherwise. The expected effects of the two productivity shocks and the coding of the dummy variables are summarized in Figure 1.4.

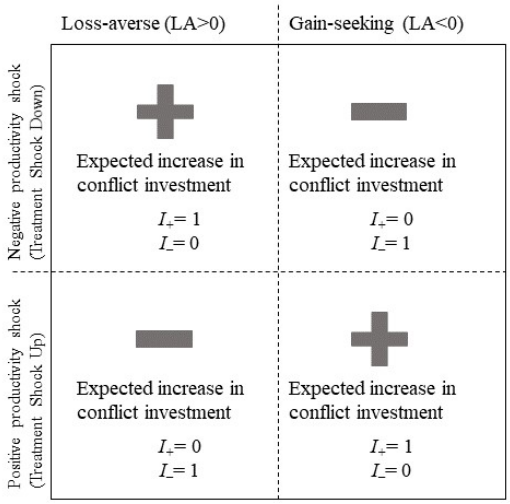


Figure 1.4. Theoretically predicted effects of positive and negative productivity shocks, conditional on the participant's loss aversion and coding of the dummy variables I_+ and I_-. The sign + (respectively −) designates a theoretically predicted positive (negative) effect of the shock. Null effect of the shock is predicted for the loss-neutral participants, thus $I_+ = I_- = 0$ for all loss-neutral participants (as well as for all participants in the Control treatment).

Table 1.2, shows a first check of the hypothesized results. The first column shows the results of regressing individual conflict investments on the interactions of the dummy variables I_+ and I_-, and the variable *Post Shock* which is set to equal one if a productivity shock has occurred prior to the round that is being played. The coefficients of the two interaction terms are to be interpreted against the baseline,

comprising of conflict investments of loss-neutral participants and of those playing in the round where no shock has occurred (which applies to all rounds in the Control treatment). Additionally, controlling for *Period Number* ensures that shock effects are not confounded by a potential time trend. The positive and significant coefficient of *Post Shock*$*I_+$ ($coef. = 2.98$, $p < 0.001$) indicates that those participants who were expected to increase their conflict investments in response to productivity shocks indeed behaved in the predicted way. However, in the group of participants who were predicted to decrease conflict investments after a shock, the experienced shocks appear not to have any significant effect ($coef. = 0.11$, $p = 0.932$).

In the second column, the control for the potential placebo effect of the shocks is added, that is, the effect of the shocks on the loss neutral participants. If a shock had an effect on its own, irrespective of loss aversion, this effect would be captured by the interaction of the variables *Post Shock* and the dummy variable *Loss Neutral* which equals one if the participant is loss-neutral ($LA = 0$), and zero otherwise. The results in column (2), however, reassuringly demonstrate that no placebo effect of the shocks is found ($coef. = 1.12$, $p = 0.576$), suggesting that those effects of the shocks that are detected and significant indeed operate through loss aversion. As the theoretical prediction suggests that there exists a positive baseline effect of loss aversion on conflict investments, in column (3), the measure of loss aversion LA is added as control. In line with the prediction of Hypothesis 3 b), the baseline effect of loss aversion is positive and significant ($coef. = 0.24$, $p = 0.006$). Finally, in column (4), demographic controls were added, but all of them had an insignificant effect. The results regarding the effect of the interactions *Post Shock*$*I_+$ and *Post Shock*$*I_-$ remain unchanged through these specifications.

In the previous analysis, participants for whom the increase (respectively decrease) in conflict investments after a productivity shock was predicted were pooled across the treatments. In order to shed more light on the dynamics behind the results, in the next step, I consider separately the data from the treatments Shock Down and Control and the data from the treatment groups Shock Up and Control. Furthermore, recalling that Hypothesis 1 and Hypothesis 2 predict an increasing effect of a productivity shock

Table 1.2. Effect of productivity shock on conflict investments - pooled results

	(1)	(2)	(3)	(4)
	Individual conflict investment: G_t^i			
Post Shock*I_+	2.98***	2.82***	2.63***	2.55***
	[0.83]	[0.89]	[0.91]	[0.91]
Post Shock*I_-	0.11	-0.05	0.02	-0.34
	[1.27]	[1.31]	[1.25]	[1.32]
Post Shock*Loss Neutral		-1.12	-0.89	-1.21
		[2.00]	[2.03]	[2.11]
LA			0.24***	0.24***
			[0.09]	[0.09]
Period Number	-0.25	-0.21	-0.19	-0.12
	[0.21]	[0.22]	[0.22]	[0.22]
Constant	18.75***	18.61***	18.60***	19.12***
	[2.52]	[2.52]	[2.51]	[2.71]
Lagged behavior	YES	YES	YES	YES
Demographics	NO	NO	NO	YES
Observations	2,778	2,778	2,778	2,640
Number of subjects	463	463	463	440

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Robust standard errors in brackets. GLS regression of individual conflict investments on dummy variables I_+ and I_- (pulling together all the participants for whom a positive (respectively negative) shock effect is predicted), interacted with the *Post Shock* dummy. All models control for the number of the round of the game and the set of variables capturing lagged conflict behavior (own and opponent's lagged conflict investment and a dummy indicating whether participant i won in the previous round). Model (2) includes additionally the control for the effect of a shock for loss-neutral participants, model (3) the measure of loss aversion and model (4) the set of demographic controls.

as the participant moves away from loss-neutrality, here, instead of simply classifying participants into either loss-averse, gain-seeking, or loss-neutral category, I also account for the degree of loss aversion or gain seeking.

Table 1.3 presents the results of regressions run only on data from treatment groups Shock Down and Control. The results depicted in the first column confirm the theoretical prediction of the effect of a negative productivity shock moderated by loss preferences. Namely, whereas the shock has no effect on its own, the interaction between the shock and the measure of loss aversion is positive and significant (*Post Shock***LA*: $coef. = -0.39$,

$p = 0.011$), indicating a positive effect of the shock on loss-averse participants ($LA > 0$), and a negative effect on gain-seeking ones ($LA < 0$). Particularly, with each step (in the sense of the previously described measures of loss preferences) away from the loss neutrality, a loss-averse (gain-seeking) participant is expected to invest around 0.4 tokens more (less) in guns, relative to the participants who have not experienced the shock or are loss-neutral. Thereby, time trend and lagged behavior are controlled for as above. Furthermore, as loss preferences are predicted to also have a baseline effect, irrespective of the shock, variable LA is additionally included as a control, and, in accordance with the theoretical prediction, it has a positive effect on conflict investments. These results are robust to the inclusion of the socio-demographic background controls (column (2)), and to the exclusion of the lagged behavior (column (3)).

Result 1. *An unexpected negative productivity shock leads to an increase in conflict investments among loss-averse participants, and a decrease in conflict investments among gain-seeking ones, relative to the conflict investments in absence of a shock. Loss-neutral participants do not change their conflict investments in response to the shock.*

The same analysis is shown in Table 1.4 for the treatments Shock Up and Control. As predicted by the theoretical analysis, the coefficient of the interaction term *Post shock*LA* is negative and (weakly) significant. However, here the positive shock appears to have a positive effect on its own, irrespective of the loss preferences. In order to shed more light on the factors driving this result, in column (2) I decompose the mediating effect of loss preferences into three variables, *Post Shock*Loss Neutral*, *Post Shock*dLA* and *Post Shock*dGS*. Thereby, *Post Shock*Loss Neutral* is an interaction of the *Post Shock* variable, with a dummy variable *Loss Neutral* indicating if the participant is loss-neutral ($LA = 0$). The other two variables capture the interaction of *Post Shock* with the participant's degree of loss aversion and gain seeking, respectively. Particularly, for any given participant, variable dLA takes the value equal to LA if the participant is loss-averse and zero otherwise. Similarly, variable dGS takes the value equal to $|LA|$ if the participant is gain-seeking and zero otherwise. Thus, these two measures are based on the distance of a given participant's loss

Table 1.3. Effect of a negative productivity shock on conflict investments

	Shock Down and Control		
	(1)	(2)	(3)
	Individual conflict investment: G_t^i		
Post Shock	-0.49	-0.51	-1.08
	[1.12]	[1.08]	[1.59]
Post Shock*LA	0.39**	0.40**	0.53**
	[0.15]	[0.16]	[0.22]
LA	0.25**	0.25**	0.45**
	[0.12]	[0.12]	[0.22]
Period number	-0.18	-0.19	1.28***
	[0.23]	[0.23]	[0.29]
Constant	22.16***	20.92***	57.07***
	[3.27]	[3.38]	[1.18]
Lagged behavior	YES	YES	NO
Demographics	NO	YES	NO
Observations	1,914	1,896	2,233
Number of Subects	319	316	319

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Robust standard errors in brackets. GLS regression of individual conflict investments on indicator variable *Post Shock*, measure of loss aversion *LA*, interactions between productivity shock and loss preferences, elicited belief of opponent's conflict investment and the set of variables capturing lagged conflict behavior (own and opponent's lagged conflict investment and a dummy indicating whether participant i won in the previous round). The regression is ran on data from treatments Shock Down and Control.

preferences from the loss neutrality ($LA = 0$), and are mutually exclusive in the sense that if one of them is positive, the other one has to equal zero. This specification allows for capturing a potentially non-linear influence of shock across the full span of loss preferences. The results depicted in column (2) indeed show that the effect of the positive productivity shock is not symmetric for loss-averse and gain-seeking participants. Namely, whereas the gain-seeking participants, in accordance with the theoretical prediction, significantly increase their conflict investments in response to the shock (*Post Shock**dGS: $coef. = 0.72$, $p = 0.001$) the shock does not decrease investments among loss-averse participants (*Post Shock**dLA: $coef. = 0.26$,

Table 1.4. Effect of a positive productivity shock on conflict investments

	Shock Up and Control			
	(1)	(2)	(3)	(4)
	Individual conflict investment: G_t^i			
Post Shock	2.48**			
	[1.06]			
Post Shock*LA	-0.27*			
	[0.16]			
Post Shock* Loss neutral		0.31	0.25	2.05
		[2.21]	[2.14]	[2.98]
Post Shock*dLA		0.26	0.26	0.53
		[0.23]	[0.24]	[0.37]
Post Shock*dGS		0.72***	0.73***	0.68**
		[0.22]	[0.22]	[0.33]
LA	0.16	0.16	0.14	0.27
	[0.12]	[0.12]	[0.12]	[0.22]
Period number	-0.20	-0.14	-0.06	1.37***
	[0.22]	[0.20]	[0.19]	[0.27]
Constant	16.89***	16.82***	23.25***	57.28***
	[2.77]	[2.76]	[3.35]	[1.42]
Lagged behavior	YES	YES	YES	NO
Demographics	NO	NO	YES	NO
Observations	1,824	1,824	1,686	2,128
Number of subjects	304	304	281	304

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Robust standard errors in brackets. GLS regression of individual conflict investments on indicator variable *Post Shock*, measure of loss aversion *LA*, interactions between productivity shock and loss preferences, elicited belief of opponent's conflict investment and the set of variables capturing lagged conflict behavior (own and opponent's lagged conflict investment and a dummy indicating whether participant i won in the previous round). The regression is ran on data from treatments Shock Up and Control.

$p = 0.26$).[11] Furthermore, allowing for the identification of shock effects conditional on the sign of loss preferences, the effect of the shock among loss-neutral participants

[11] Applying this specification in the case of a negative productivity shock does not substantially change the results reported in Table 1.3. Particularly, substituting variable *Post Shock*LA* by the interaction terms *Post Shock*Loss Neutral*, *Post Shock*dLA* and *Post Shock*dGS* in column (1) results again in: no significant effect of the shock on loss-neutral participants (*Post Shock*Loss Neutral*: $coef. = -2.33$, $p = 0.462$), positive effect among loss-averse participants (*Post Shock*dLA*: $coef. = 0.36$, $p = 0.071$), and negative effect among gain-seeking participants (*Post Shock*dGS*: $coef. = -0.39$, $p = 0.076$)

becomes small and insignificant, as expected.[12] As above, these results are robust to the inclusion of socio-demographic controls (column (3)), and to the exclusion of lagged behavior (column (4)).[13]

Result 2. *An unexpected positive productivity shock leads to an increase in conflict investments among gain-seeking participants, but does not cause the loss-averse participants to decrease their conflict investments. Loss-neutral participants do not change their conflict investments in response to the shock.*

1.4.3 Loss aversion in absence of productivity shock

Hypothesis 3 states that the optimal conflict investment is independent of the productivity level and increases in loss aversion in the case where we can regard the reference lottery to be equal to the lottery that is actually played, that is, in the absence of a productivity shock. Conflict investments are predicted to increase in loss aversion across the whole range of loss preferences values, that is, the more loss-averse, or equivalently the less gain-seeking the player is. The first test of this hypothesis was already provided in Table 1.2, where in accordance with the theoretical prediction, the measure of loss aversion (LA) was shown to have a positive and significant effect on conflict investment while controlling for the effects of the shock. Table 5 presents a stricter test of this effect.

In order to isolate only "the baseline effect of loss aversion" (absent any productivity shock), here the regression was run only on the data from rounds where no shock previously took place.[14] The dependent variable remains conflict investment from the current period (G_t^i), and all controls are as explained above.

[12] As a robustness check, the effect of the positive shock among loss-neutral participants could also be measured by changing the specification in column (2) such that the interaction term *Post Shock*Loss Neutral* is substituted by the variable *Post Shock* (while retaining the other two interaction terms as controls). The results reported in Table 1.4 are not qualitatively changed with this specification (*Post Shock*: $coef. = 1.03$, $p = 0.472$; *Post Shock*dLA*: $coef. = 0.10$, $p = 0.69$; *Post Shock*dGS*: $coef. = 0.62$, $p = 0.011$)

[13] Simple OLS estimations of models in Table 1.3 and 1.4 confirming the results reported in this section are reported in Tables 1.8 and 1.9 in Appendix.

[14] That is, only data from the rounds 2 through 4 from treatment groups Shock Down and Shock Up and from rounds 2 through 7 in the Control group are taken into account.

Table 1.5. Baseline effect of loss aversion on conflict investments

	(1)	(2)
	Individual conflict investment: G_t^i	
LA	0.20*	0.20*
	[0.11]	[0.11]
Productivity	-0.22	-0.22
	[0.21]	[0.20]
Period number	-0.14	-0.10
	[0.23]	[0.23]
Constant	18.65***	21.57***
	[3.48]	[3.87]
Lagged behavior	YES	YES
Demographics	NO	YES
Observations	1,869	1,791
Number of subjects	463	440

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Robust standard errors in brackets. GLS regression of individual conflict investments on measure of loss aversion, productivity level, period number, set of variables capturing lagged conflict behavior (own and opponent's lagged conflict investment and a dummy indicating whether participant i won in the previous round) and a set of demographic variables. In order to isolate the baseline effect of loss aversion in absence of a productivity shock, the regression was ran only the data from the rounds where no shock took place.

The results shown in the first column show that the coefficient of loss aversion is positive, following the theoretical prediction, although now with reduced significance ($coef. = 0.20$, $p = 0.065$). Furthermore, the productivity level, as expected, has no significant impact on conflict investments. The results in column (2) demonstrate that including demographic controls does not qualitatively change the results.

Result 3. *Loss aversion in absence of a productivity shock has a positive effect on conflict investments.*

1.4.4 Alternative mechanisms - joy of winning and cognitive reflection

One possible alternative explanation for the observed effect of the productivity change on conflict investments could be the joy of winning – the utility that agents derive from winning per se, independent of associated payoff. Joy of winning has already been considered as a potential source of overbidding and heterogeneity of investments in contest in multiple experimental works (see for example Cason et al. (2018), Herbst et al. (2015), Konrad and Morath (2019), Price and Sheremeta (2011), Price and Sheremeta (2015), Sheremeta (2010)), and has often been found to increase the contest efforts.

In order to analyze its influence in the context of a conflict game, consider again the guns-and-butter game, where the marginal return of the productive activity is set at β, but now allowing for the possibility that additional to the monetary value of the prize won, the participants also have an additive non-monetary constant utility of winning. Denoting this additional utility of winning by w (and assuming that w is common for both players), the expected utility of player i can be represented as:

$$E[V(G_i, G_j)] = \frac{G_i}{G_i + G_j}\left(\beta\left(2E - G_i - G_j\right) + w\right)$$

The unique Nash equilibrium, denoted by G_w^*, equals:

$$G_i^* = G_j^* = G_w^* = \frac{E}{2} + \frac{w}{4\beta}$$

Thus, a higher value of joy of winning (w) is expected to lead to a higher optimal conflict investment for any given productivity level. Furthermore, we would expect that, for any given positive value of joy of winning, optimal conflict investment decreases in productivity level. Therefore, a negative (positive) shock in productivity could lead to an increase (decrease) in conflict investments due to a positive joy of winning rather than as a consequence of loss aversion.

In order to test this result, a measure of joy of winning was elicited using the test developed by Sheremeta (2010). After completing the guns-and-butter game, participants played one round of the Tullock contest for the zero value prize. For this part of the experiment, the participants were given an additional endowment of 100 tokens, out of which they could invest any integer number of tokens in the zero prize contest. In the observed sample, 57% of the participants invested a positive number of tokens in this task and the median value of investments equals 10 tokens.

Table 1.6 presents the results of the regression of conflict investments, run separately on data from treatment groups Shock Down and Control (columns (1) and (2)), and on treatment groups Shock Up and Control (columns (3) and (4)). Additionally to the variables discussed above, the elicited measure of joy of winning, which was taken to be equal to the investment in the zero prize Tullock contest (JoW), as well as its interaction with productivity ($Post\ Shock^*JoW$) are also included.

The results shown in column (1) and column (3) of Table 1.6, demonstrate that including joy of winning in the regression does not change the earlier found effect of productivity shocks operating through loss aversion. Specifically, considering the data of the Shock Down and Control treatments only (column (1)), joy of winning has a negative effect on conflict investments, contrary to the theoretical prediction ($coef. = -0.04$, $p = 0.011$). Observing the data from the treatments Shock Up and Control (column (3)), joy of winning had no significant effect. Furthermore, the coefficient of the interaction $Post\ Shock^*JoW$ in columns (1) and (3), shows that joy of winning had no role in determining the reaction to either positive or negative productivity shock. Thereby, the effects of both positive and negative productivity shocks remain unchanged to what was found above.[15]

[15]The results regarding the impact of joy of winning should however be taken with caution, as the distribution of JoW appears to differ among the treatments. The average JoW in the Shock Down treatment equals 34.55, whereas the average values in Shock Up and Control treatments are 28.18 and 27.78, respectively. Further, while the distributions of JoW in the treatments Shock Up and Control do not differ from each other (Mann-Whitney test, $z = 0.1$, $p = 0.92$), the distribution of JoW in the Shock Up treatment differs when compared to either of the two other treatments, although this difference is not significant (Mann-Whitney test comparing Shock Down and Shock Up, $z = 1.8$, $p = 0.072$; Mann-Whitney test comparing Shock Down and Control, $z = -1.7$, $p = 0.089$). This indicates that as the zero-prize lottery task was administered directly after playing the seven rounds of the guns-and-butter game, and as the tasks are relatively similar, the measured value of joy of winning might be influenced by the treatment effects.

Table 1.6. Effects of joy of winning and cognitive reflection on conflict investments

	(1)	(2)	(3)	(4)
	Shock Down and Control		Shock Up and Control	
	Individual conflict investment: G_t^i			
JoW	-0.04**	-0.04**	-0.01	-0.01
	[0.01]	[0.01]	[0.01]	[0.01]
Post Shock*JoW	0.00	0.00	0.01	0.01
	[0.02]	[0.02]	[0.02]	[0.02]
CRT		-0.03		0.01
		[0.36]		[0.34]
Post Shock*Loss neutral	-1.63	-1.62	0.16	0.16
	[3.43]	[3.44]	[2.12]	[2.11]
Post Shock*dLA	0.39*	0.39*	0.23	0.23
	[0.22]	[0.22]	[0.23]	[0.23]
Post Shock*dGS	-0.39*	-0.39*	0.70***	0.70***
	[0.23]	[0.23]	[0.21]	[0.22]
LA	0.22*	0.23*	0.15	0.15
	[0.12]	[0.13]	[0.12]	[0.12]
Period number	-0.23	-0.23	-0.15	-0.15
	[0.23]	[0.23]	[0.20]	[0.20]
Constant	23.86***	23.97***	17.06***	17.02***
	[3.35]	[3.82]	[2.75]	[2.89]
Lagged behavior	YES	YES	YES	YES
Observations	1,914	1,914	1,824	1,824
Number of subjects	319	319	304	304

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Robust standard errors in brackets. GLS regression of individual conflict investments on elicited measure of joy of winning (JoW), the interaction of joy of winning and productivity level, interaction of the degree of loss aversion (gain seeking) with the dummy indicating whether productivity shock took place in one of the previous rounds, measure of loss aversion and CRT score. The left part of the table (columns (1) and (2)) include only data from the "Shock Down" treatment and the control group. The right part of the table (columns (3) and (4)) include only data from the treatment "Shock Up" and the control group.

Finally, the results shown in columns (2) and (4) demonstrate that the effect of the score in the cognitive reflection test on conflict investments was highly insignificant, and its inclusion led to virtually no change in the coefficients of other variables and their significance. It thus does not appear that ability to understand the mechanics of the

guns-and-butter game, as proxied by the cognitive reflection score, had an influence on the conflict behavior.

1.4.5 Escalation in conflict investments

The work of Lacomba et al. (2014) studies the time development of conflict investments and the phenomenon of conflict escalation in the guns-and-butter game.[16] In this subsection, I revert to the analysis of conflict investments development over the rounds of the game in my sample and compare my results to their findings. In order to avoid the confounding effects of the productivity shocks on conflict investments, in the following analysis, I consider only the data from the rounds where no productivity shock yet took place.

As stated above, the average investment in conflict is significantly higher than the one predicted by the Nash equilibrium with rational risk and loss neutral players (G_{NE}) in all three treatments. Contrary to the results of Lacomba et al. (2014), who find that the average conflict investment at the inception of the game does not differ from G_{NE} and escalate only over time, I find that the average conflict investment significantly exceeds this level already in the first round of the game (in all three treatments WSR test, $p < 0.001$). Thus, even when any effect of the repeated game structure investments from previous rounds) is isolated-away, the investment in conflict is still substantially higher than the theoretically predicted level G_{NE}. Furthermore, the average difference to G_{NE} further increases over rounds of the game, corroborating Lacomba's conflict escalation result.

Following the approach of Lacomba et al. (2014), I first calculate the time trend in all three treatments by calculating Spearman's rank correlation coefficient between the conflict investments and period number for each individual (ρ^i). Thereafter, I use a non-parametric test to check whether the median of time trend differs significantly from zero. Lacomba et al. (2014) find a significantly positive time trend, but only if the plain

[16] One of the treatments in Lacomba et al. (2014), employs the same game and exogenously sets the take-rate to 100% (the share of the loser's production that the winner appropriates), and thus corresponds exactly to the game played here (at least in the first four rounds, before any productivity change is introduced).

guns-and-butter game is complemented by the possibility of the winner of the conflict to decide on how much to appropriate from the loser. Contrary to their results, suggesting an important role of the ability to decide the degree of expropriation for the escalation of conflict, I find a positive time trend in my sample (average ρ^i equals 0.09), although here the degree of expropriation by the winner is exogenously set to one. The time trend significantly differs from zero, both if the sample is considered as a whole (WSR test, $p < 0.001$) and if the treatments are considered separately (in all three treatments WSR test, $p < 0.001$).

In order to shed more light on the mechanism behind this escalation, Table 6 shows results of the regression of conflict investments on the period number in model (1), and then adding the previous round difference between own and opponent's conflict investments, and the dummy indicating whether a participant lost the conflict in the previous round (*i lost*) in model (2). The difference between the own and opponent's conflict investments from the previous round is broken into two variables. $\left|G^i_{t-1} - G^j_{t-1}\right|_+$ equals the absolute value of a difference in investments if player i invested more in guns in the previous round than their opponent from that round, and zero otherwise. Similarly, $\left|G^i_{t-1} - G^j_{t-1}\right|_-$ equals the absolute value of a difference in investments if player i invested less than their opponent in guns in the previous round, and zero otherwise. This should enable detecting a potentially asymmetrical reaction to the observed difference in conflict investments in cases where a participant is leading, compared to when they are lagging behind their opponent's investments.

As the results demonstrate, the coefficient of the *Period Number* goes from highly significant and positive in the first model, to close to zero and insignificant if the controls for past conflict behavior are added in the next two models. Thus rather than the passage of time per se, it is the dynamics of reaction to the lagged own and opponents' past conflict investments, as well as the effect of having won or lost in the previous round that explain the conflict escalation.

The difference between lagged own and opponent's conflict investments has a significant negative effect if own investment was higher than opponent's ($coef. = 0.05$, $p = 0.06$), and conversely has a significant positive effect if own investment was lower than that

Table 1.7. Conflict investments escalation over time

	(1)	(2)	(3)
	Individual conflict investment: G_t^i		
Period Number	0.88***	-0.08	-0.33
	[0.31]	[0.22]	[0.21]
G_{t-1}^i		0.70***	0.55***
		[0.03]	[0.05]
$\left\|G_{t-1}^i - G_{t-1}^j\right\|_+$		-0.05*	0.02
		[0.03]	[0.03]
$\left\|G_{t-1}^i - G_{t-1}^j\right\|_-$		0.20***	0.12***
		[0.05]	[0.04]
i lost		2.52***	2.35***
		[0.85]	[0.83]
Opponent's guns guess			0.30***
			[0.04]
Constant	59.42***	16.88***	9.08***
	[1.48]	[2.65]	[2.47]
Observations	1,869	1,869	1,869
Number of subjects	463	463	463

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Robust standard errors in brackets. GLS regression of individual conflict investments on period number, positive and negative (absolute) difference between own and opponent's conflict investments in the previous round, a dummy indicating whether the participant lost the conflict in the previous round. Model (3) additionally includes the elicited belief of opponent's conflict investment.

of the opponent in the previous period ($coef. = 0.20$, $p < 0.001$). However, the latter coefficient is four times higher (in absolute terms) than the former, meaning that the increase in conflict in reaction to "lagging behind" outweighs the decrease when "leading", and creates an upward trend. In accordance with the results of Lacomba et al. (2014), the effect of losing in the previous round has an additional positive and significant effect, as the losers increase their conflict investments in the following round.

Furthermore, including the elicited belief of the opponent's conflict investment, denoted by *Opponent's guns guess*, as control (elicited prior to each round of the game) shows that

the belief has a positive and significant effect on own conflict investments.[17] Furthermore, the results depicted in model (3) demonstrate that the effect of lagged behavior partially functions through belief formation, as after including *Opponent's guns guess* the effect of $\left|G^i_{t-1} - G^j_{t-1}\right|_+$ becomes insignificant, and the effect of $\left|G^i_{t-1} - G^j_{t-1}\right|_-$ although still significant, is almost halved.

1.5 Conclusion

This work provides an analysis of the behavior in the guns-and-butter conflict game played under changing productivity regimes. In doing so, it provides an alternative and complementary explanation for conflict behavior in the aftermath of an income shock. Rather than extending the model to include asymmetries in shock effects for different sectors of the economy, or considerations of future economic recovery, as previously proposed in the literature, I allow only for the inclusion of expectation based loss preferences while leaving all other elements of the model as simple as possible.

The theoretical discussion shows that loss aversion, that is, attaching higher (absolute) utility to the losses than to the gains of similar size, is sufficient to explain higher levels of conflict after a negative productivity shock (and lower conflict levels after a positive one). On the other hand, gain seeking, which implies showing higher sensitivity to gains than to the equal-sized losses, is predicted to lead to a decrease in conflict investments in the aftermath of a negative productivity shock, and to an increase after a positive one. Furthermore, in the absence of any productivity change, loss aversion is expected to increase conflict investments.

The results of the conducted lab experiment show that, in accordance with theoretical predictions, after a negative productivity shock, loss-averse individuals increase their conflict investments, and gain-seeking participants decrease them. As loss-averse (rather than gain-seeking) individuals are expected to constitute a majority, as is the case in

[17] This is remarkable considering that the best response of player i to the conflict investment of player j (assuming a perfectly rational risk- and loss-neutral player) can be expressed as $G_i^{BR}(G_j) = \sqrt{2EG_j} - G_j$, and thus decreases in G_j for all $G_j > 50$ tokens. Considering that over 70% of all guesses were higher than 50 tokens, if the participants behaved according to the theoretical prediction, we would expect to see a negative coefficient of *Opponent's guns guess*, rather than the significant positive effect that we observe.

the observed sample, this provides an additional, preference-based explanation for the effects of adverse shocks in income on conflict behavior found in the empirical literature. However, while positive productivity shock leads to the predicted increase in conflict investments among gain-seeking participants, it fails to reduce conflict investments among loss-averse ones. Finally, independent of productivity level and in the absence of any productivity changes, a higher degree of loss aversion is associated with higher conflict investments.

A potential effect of the joy of winning and that of participants' cognitive abilities are additionally considered, as they might provide an alternative explanation of the productivity shock effects. Experimental results demonstrate that, while joy of winning had the expected effect on conflict investments in the treatment where a negative productivity shock was administered, it had no effect on the behavior in the treatment where a positive shock took place. Cognitive ability did not have any impact on participants' conflict behavior, providing evidence against attributing changes in conflict behavior in the aftermath of a shock to the lack of understanding of the mechanics of the guns-and-butter game. Moreover, controlling for these behavioral traits of participants did not change the previously found results regarding the role of loss aversion, corroborating the evidence for the loss preferences indeed being a channel through which unexpected productivity changes affect participants' conflict investments.

1.6 Appendix

1.6.1 Alternative specifications

Tables 1.8 and 1.9 report the results of OLS regressions analogous to the GLS estimates reported in Table 1.3 and Table 1.4. The results discussed there are here confirmed.

Table 1.8. Effect of negative productivity shock on conflict investments

	Shock Down and Control			
	(1)	(2)	(3)	(4)
	Individual conflict investment: G_t^i			
Post Shock*dLA	0.60***	0.39*	0.36*	0.52***
	[0.19]	[0.20]	[0.20]	[0.19]
Post Shock*dGS	-0.64***	-0.37	-0.39*	-0.44*
	[0.19]	[0.22]	[0.22]	[0.22]
LA		0.25**	0.25*	0.11
		[0.12]	[0.12]	[0.11]
Post Shock*Loss neutral			-2.33	-1.61
			[3.17]	[3.09]
Opponent's guns guess				0.30***
				[0.04]
Period Number	-0.26	-0.24	-0.20	-0.42*
	[0.22]	[0.22]	[0.22]	[0.21]
Constant	22.42***	22.34***	22.14***	13.23***
	[3.22]	[3.22]	[3.23]	[3.24]
Lagged behavior	YES	YES	YES	YES
Observations	1,914	1,914	1,914	1,914
R-squared	0.33	0.33	0.33	0.38

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Robust standard errors in brackets. OLS regression of individual conflict investments on indicator variable *Post Shock*, measure of loss aversion LA, interactions between productivity shock and loss preferences, elicited belief of opponent's conflict investment and the set of variables capturing lagged conflict behavior (own and opponent's lagged conflict investment and a dummy indicating whether participant i won in the previous round). The regression is ran on data from treatments Shock Down and Control.

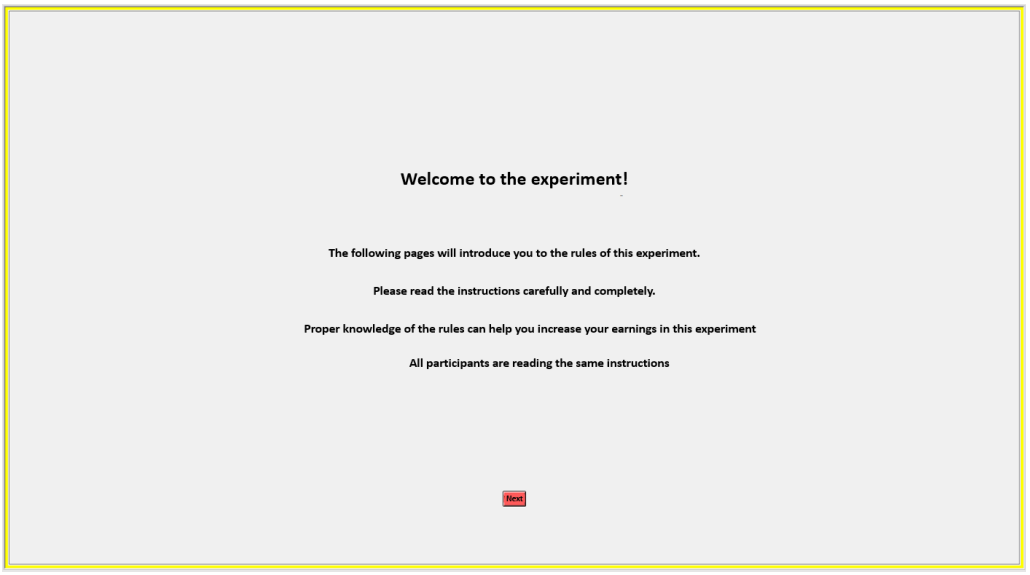
Table 1.9. Effect of positive productivity shock on conflict investments

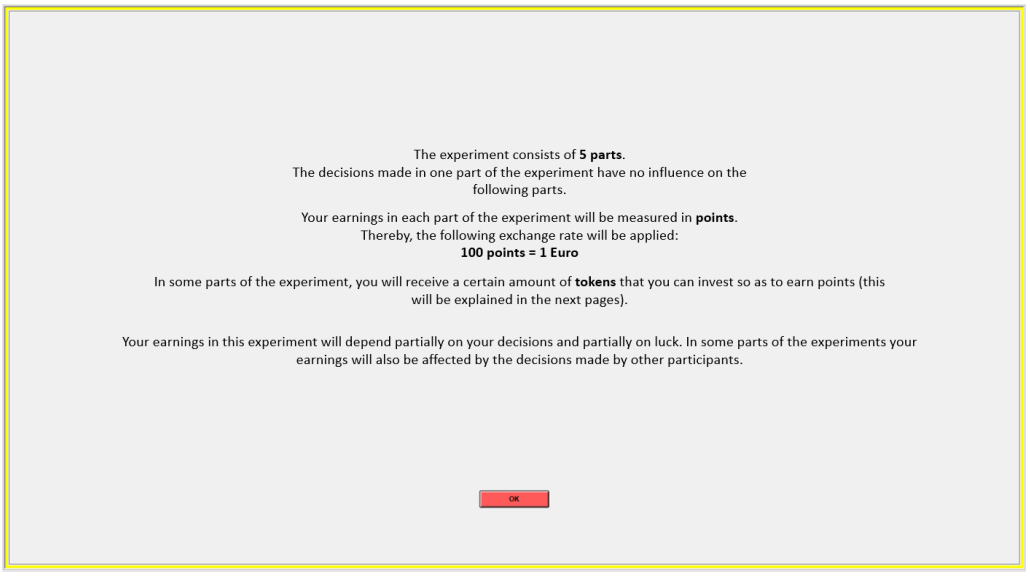
	Shock Up and Control			
	(1)	(2)	(3)	(4)
	Individual conflict investment: G_t^i			
Post Shock*dLA	0.38*	0.25	0.26	0.32
	[0.22]	[0.22]	[0.23]	[0.20]
Post Shock*dGS	0.55***	0.72***	0.72***	0.69***
	[0.18]	[0.21]	[0.22]	[0.21]
LA		0.16	0.16	0.08
		[0.12]	[0.12]	[0.11]
Post Shock*Loss neutral			0.31	-1.36
			[2.21]	[2.13]
Opponent's guns guess				0.28***
				[0.04]
Period Number	-0.16	-0.14	-0.14	-0.46**
	[0.20]	[0.20]	[0.20]	[0.21]
Constant	16.88***	16.81***	16.82***	10.39***
	[2.76]	[2.76]	[2.76]	[2.48]
Lagged behavior	YES	YES	YES	YES
Observations	1,824	1,824	1,824	1,824
R-squared	0.34	0.34	0.34	0.38

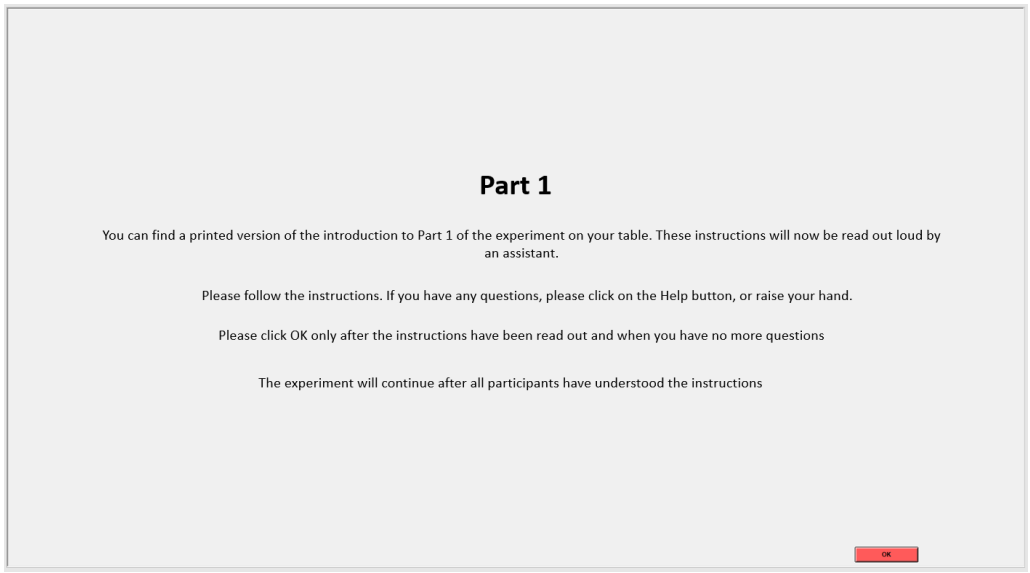
Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Robust standard errors in brackets. OLS regression of individual conflict investments on indicator variable *Post Shock*, measure of loss aversion *LA*, interactions between productivity shock and loss preferences, elicited belief of opponent's conflict investment and the set of variables capturing lagged conflict behavior (own and opponent's lagged conflict investment and a dummy indicating whether participant i won in the previous round). The regression is ran on data from treatments Shock Up and Control.

1.6.2 Instructions

The instructions shown to participants for each part of the experiment are shown below. The instructions shown here are the translations into English of the original instructions in German language shown to participants. The instructions for the guns-and-butter game shown here, are the instructions for the treatment Shock Up. For the treatment groups Shock Down and Control, the instructions remain the same in formulation, and only the productivity levels are changed accordingly.







Instructions for Part 1:

Players

In each round of Part 1 you will be matched into a pair with one other participant. In each new round you will be matched with another person, thus you can never play with the same person in more than one round. For simplicity we will sometimes refer to the other player with whom you are paired in one round as "Other". At the beginning of each round the computer will randomly assign you to be called either "**Player A**" or "**Player B**".

Earnings

In Part 1 of the experiment you will be given some Tokens that you can use to earn Points (recall that the total points that you earn in the experiment will be converted into euros at the end of the experiment at rate: 100 Points = 1 EUR). Particularly, at the beginning of **each round** both you and the participant that you are paired with (Other) will be given an amount of 100 Tokens to earn points with. At the end of this phase of the experiment, **one of the rounds will be randomly selected by the computer to be paid out**. The earnings of that round, together with earnings from other phases of the experiment and the show-up fee, will then be paid out in private at the end of the experiment.

Rules

In each round of the game, you as well as the Other will have to allocate the 100 Tokens that each of you has received to two projects: **project P1** (we will call an investment in project P1 "P1 Tokens") and **project P2** (we will call investment in project P2 "P2 Tokens"). Any distribution is allowed, including putting all tokens in only one project. P1 Tokens and P2 Tokens lead to possibility of earning points as described:

Project P1:

For each Token put into P1 it holds that: **1 Token** = **6 Points**
Thus for each Token put into P1 you generate **6 Points**.

Project P2:

This project concerns a lottery. The Tokens that you and the Other put into P2 will determine your and Other's chances of winning this lottery. Your chance to win the lottery is determined by your share in the total number of Tokens invested in P2:
Your chance of winning = Your P2 Tokens / (Your P2 Tokens + Other's P2 Tokens)

Similarly, Other's chance to win the lottery is determined by Other's share in the total number of P2 Tokens. Thus, your chance of winning and Other's chance of winning sum up to 100%.
After you and the Other have made decisions on how many Tokens you want to invest in P2, which determines your probabilities of winning, the computer will randomly determine the winner of the lottery taking these winning probabilities into account (via wheel of fortune).

The winner of this lottery receives **all the points that were generated in project P1 by both players**, and **the loser receives zero points**. This is to say that if you win the lottery you receive:
all the points that you generated in project P1 **PLUS** all the points that Other generated in project P1.
If you lose the lottery you receive 0 points.

After everyone has read the instructions you will be shown some examples on the screen that should help you to familiarize with the rules. You will also be asked some questions regarding the rules.

The instructions for the other parts of the experiment will be provided at the beginning of those parts.

Example 1:
At the beginning of each round, two participants will be assigned to a group, and each of them will receive 100 tokens. In the following step, each participant has to decide how to divide their 100 tokens between investment project P1 and investment project P2. Imagine that the two participants decided to split their tokens between the two projects as shown below (please note that the investments shown here are hypothetical and were randomly generated by the computer.)

Participant A	Partcipant B
Investment in P1: 12 Tokens	**Investment in P1: 31 Tokens**
- *Generates 12 x 6 = 72 Points*	- *Generates 31 x 6 = 186 Points*
Investment in P2: 88 Tokens	**Investment in P2: 69 Tokens**

In total, 258 points were generated in investment project P1 (participant A generated 72 points and participant B generated 186 points).
Therefore, **the winner will get 258 points** and **the looser gets 0 points.**

A total of 157 tokens were invested in investment project P2 (participant A invested 88 tokens and participant B invested 69 tokens). The winning probabilities are based on the investments in project P2:
Winning probability of participant A = 88/(88+69)=88/157=56.1%
Winning probability of participant B = 69/(88+69)=69/157=44.0%

Based on those winning probabilities, the computer creates a wheel of fortune, which determines the winner (see the figure below). The area marked in red, which represents 56.1% of the wheel's area, represents the winning probability of participant A. The area marked in blue, which represents 44.0% of the wheel's area, represents the winning probability of participant B. In the next step, the arrow of the wheel will be spinning, and depending on where it stops the winner will be determined.

OK

Example 2:
At the beginning of each round, two participants will be assigned to a group, and each of them will receive 100 tokens. In the following step, each participant has to decide how to divide their 100 tokens between investment project P1 and investment project P2. Imagine that the two participants decided to split their tokens between the two projects as shown below (please note that the investments shown here are hypothetical and were randomly generated by the computer.)

Participant A	Partcipant B
Investment in P1: 21 Tokens	**Investment in P1: 21 Tokens**
- *Generates 21 x 6 = 126 Points*	- *Generates 21 x 6 = 126 Points*
Investment in P2: 79 Tokens	**Investment in P2: 79 Tokens**

In total, 252 points were generated in investment project P1 (participant A generated 126 points and participant B generated 126 points).
Therefore, **the winner will get 252 points** and **the looser gets 0 points.**

A total of 157 tokens were invested in investment project P2 (participant A invested 88 tokens and participant B invested 69 tokens). The winning probabilities are based on the investments in project P2:
Winning probability of participant A = 79/(79+79) = 79/158 = 50.0%
Winning probability of participant B = 79/(79+79) = 79/158 = 50.0%

Based on those winning probabilities, the computer creates a wheel of fortune, which determines the winner (see the figure below). The area marked in red, which represents 50.0% of the wheel's area, represents the winning probability of participant A. The area marked in blue, which represents 50.0% of the wheel's area, represents the winning probability of participant B. In the next step, the arrow of the wheel will be spinning, and depending on where it stops the winner will be determined.

OK

Test question 1 from 4

Imagine the following scenario:
At the beginning of each round, two participants will be assigned to a group, and each of them will receive 100 tokens. In the following step, each participant has to decide how to divide their 100 tokens between investment project P1 and investment project P2. Imagine that the two participants decided to split their tokens between the two projects as shown below.
Note that each token invested in project P1 generates **6 points**. You can use the calculator to answer this question (please click on the "calculator" button).

Participant A	Partcipant B
Investment in P1: 80 Tokens	Investment in P1: 50 Tokens
Investment in P2: 20 Tokens	Investment in P2: 50 Tokens

The winner will get as a prize all the points that were generated in project P1 from investments of both players.
How many points will the winner receive in this example?

Check

Test question 2 from 4

The investments of both participants remain the same as on previous screen (see below). What are the winning probabilities of both participants?

Participant A	Partcipant B
Investment in P1: 80 Tokens	Investment in P1: 50 Tokens
Investment in P2: 20 Tokens	Investment in P2: 50 Tokens

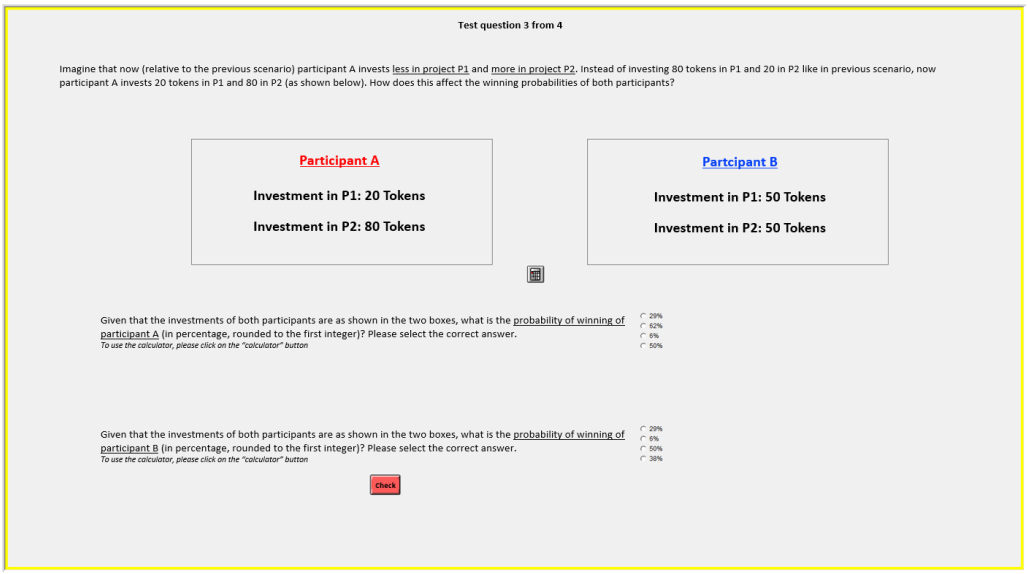
Given that the investments of both participants are as shown in the two boxes, what is the probability of winning of participant A (in percentage, rounded to the first integer)? Please select the correct answer.
To use the calculator, please click on the "calculator" button

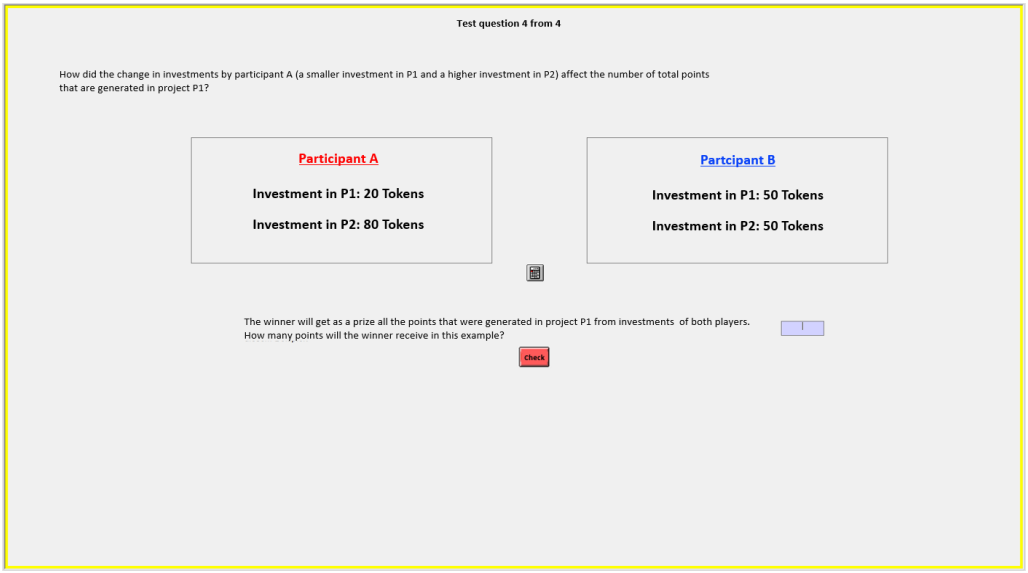
- 29%
- 9%
- 23%
- 15%

Given that the investments of both participants are as shown in the two boxes, what is the probability of winning of participant B (in percentage, rounded to the first integer)? Please select the correct answer.
To use the calculator, please click on the "calculator" button

- 15%
- 71%
- 9%
- 23%

Check





After the first four rounds of the game, the participants were informed about the change in productivity, as shown on the next screen-shot below.

Rounds 5 to 7:

In rounds 5 to 7, each token invested in project P1 generates 10 points, instead of 6 as was the case in previous rounds.

All other rules apply as described.

OK

Part 2

Instructions:

In Part 2 you need to make a single investment decision. The rules are similar as in Part 1 but the following changes to the rules apply: after you are assigned to a group with one other participant (this can be any participant in the lab) and after you are randomly assigned the role of "participant A" or "participant B", you will again receive 100 tokens. Differently than in Part 1, where you had to invest in two projects, here you can only invest in one project – project P3.

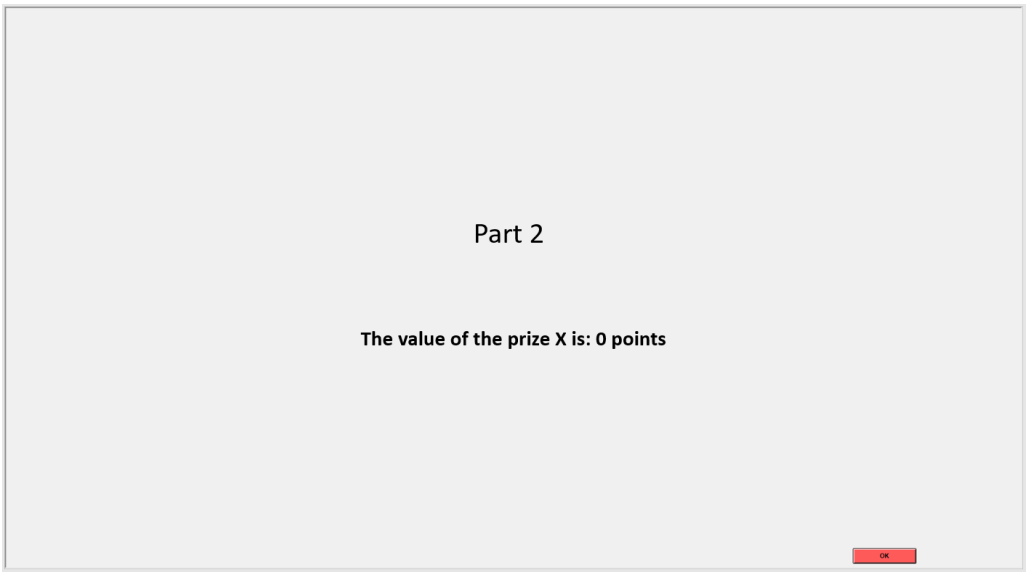
The investment decisions of you and the other participant with whom you were matched will determine your probabilities of winning the fixed prize in value of X points. Your winning probability equals your share of P3 tokens in the total P3 tokens:

Your winning probability = Your P3 tokens / (Your P3 tokens + P3 tokens of the other participant)

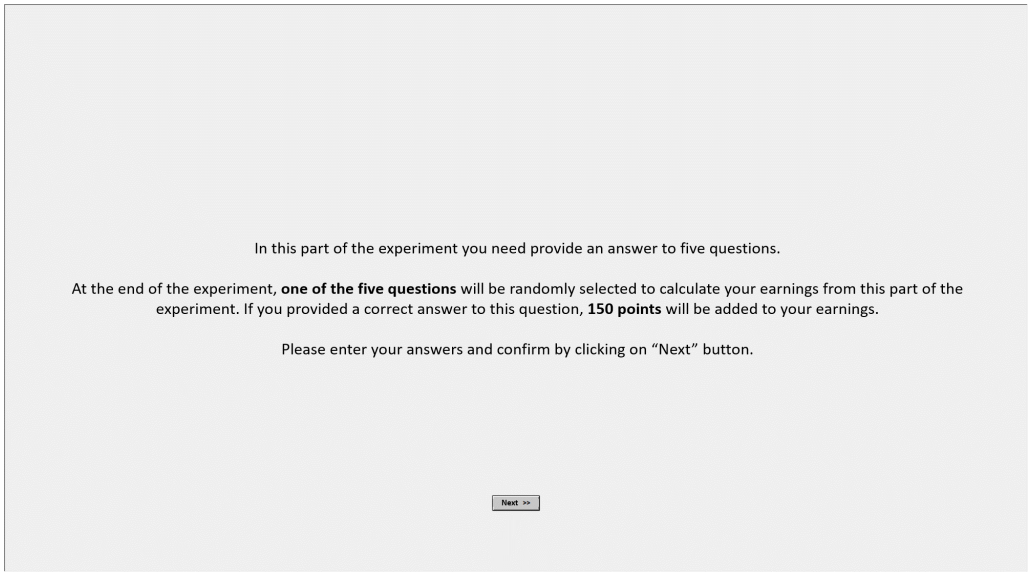
You can invest as many tokens in project P3 as you wish (0 to 100 tokens). All tokens that you do not invest will be exchanged to points at rate 1 token = 1 point, and those points will be paid out to you at the end of the experiment.

After you have made your investment decision, the computer will decide the winner via wheel of fortune, in the same way as in the previous rounds. The winner receives the prize in value of X points. The exact value of the prize, X will be disclosed to you on the next screen.
The looser receives 0 points.

OK

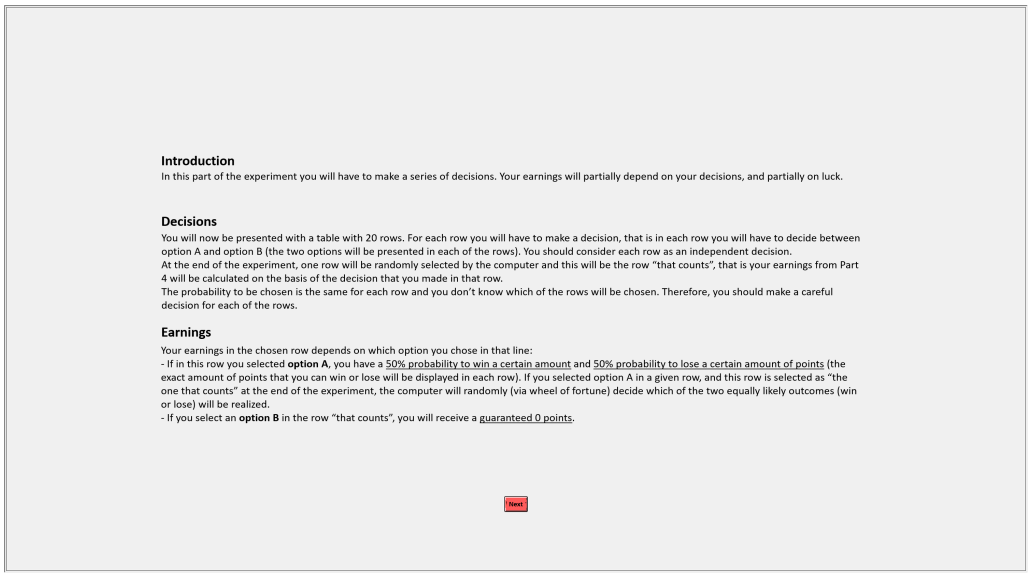

Part 2
The value of the prize X is: 0 points
OK


Part 3
An introduction to the Part 3 of the experiment will be displayed on the next screen.
Please read the instructions carefully and completely.
Next



The questions used to assess cognitive reflection are stated below:

- A bat and a ball cost \$1.10 in total. The bat costs a dollar more than the ball. How much does the ball cost?
- If it takes 2 nurses 2 minutes to measure the blood pressure of 2 patients, how many minutes would 200 nurses need in order to measure the blood pressure of 200 patients?
- In a lake, there is a patch of lily pads. Every day, the patch doubles in size. If it takes 48 days for the patch to cover the entire lake, how long would it take for the patch to cover half of the lake?
- Max received both the 15th highest and the 15th lowest mark in the class. How many students are in the class?
- Simon decided to invest \$8,000 in the stock market one day early in 2008. Six months after he invested, on July 17, the stocks he had purchased were down 50%. Fortunately for Simon, from July 17 to October 17, the stocks he had purchased went up 75%. At this point: a) the value of purchased shares is equal to their value at the time of the purchase, b) the value of purchased shares is higher than their value at the time of the purchase, c) the value of purchased shares is lower than their value at the time of the purchase.



Participants were then presented with the table as depicted in Table 1, were in each of the rows they could select either option A or option B.

Part 5

In this part of the experiment you will be asked to provide answers to a couple of questions.

You will be rewarded by 500 points for answering these questions completely.

Next

Questionnaire

Before you get your final earnings, please answer the following questions:

How old are you?

Gender
- Female
- Male

Is German your mother tongue
- Yes, German is my mother tongue.
- No, German is not my mother tongue.

What is the field of your studies?
- Economics / Business Administration
- Social Sciences
- Humanities
- Law
- Medicine / Veterinary Medicine
- Natural Sciences
- Mathematics, Computer Science, Engineering
- Psychology, Pedagogy
- Other field of study
- Not a student

How many other participants in this session do you know, at least superficially?
- None
- One participant
- Two participants
- More than two participants

Do you know somebody who already participated in this experiment?
- No, nobody
- Yes, one person
- Yes, more persons

Next

Chapter 2

Queen Bee Immigrant:

The effects of status perceptions on immigration attitudes

2.1 Introduction

The indications that Alternative für Deutschland, an Euro-sceptic right-wing party in Germany that based the core of its platform on opposing immigration, had reached higher electoral support in the 2017 federal election among the so-called Russian-speaking German community compared to the national average (Goerres et al. (2020)), attracted a lot of media attention in Germany. Indeed, this is seemingly counter-intuitive – why would groups who themselves have a history of immigration and are also largely perceived by natives as immigrants support anti-immigration platforms? This is, however, not a sole example of such inter-minority dynamics. Cases of negative immigration attitudes expressed by the groups of immigrants were also found, for example, in Switzerland (Strijbis and Polavieja (2018)), Belgium (Meeusen et al. (2019)) and Austria (Neuhold (2020)).

This work focuses on studying the dynamics of inter-minority relations and attempts to uncover the influence of the minority-group's status position in the host country on it's members' attitudes towards other minorities. I hypothesize that relative status deprivation, that is, the negative difference in status between own ethnic/national group

and that of the native majority (or other, more favorably perceived minorities), has a negative impact on group's members' attitudes toward an even lower ranked status group (e.g., refugees).

While a considerable body of scientific literature studies the attitudes of the majority population toward migration (for a survey of this literature see, e.g., Hainmueller and Hopkins (2014)), less attention is paid to the political and immigration attitudes of established migrants and their determinants. In principle, factors as diverse as those that have been found to impact the immigration attitudes of the majority population, including shared cultural values and perceived economic or cultural threat, could be determining the positions of established immigrants as well. Furthermore, prevailing socio-economic conditions in a given immigrant group might impact its members' views, as for example could be the case if the group is over-represented in an employment sector that is perceived to be particularly affected by the inflow of new immigrants. Additionally, cultural and political characteristics of the sending country, including potential histories of conflicts with certain national or religious groups, might as well influence the attitudes of individuals immigrating from that country. Notwithstanding the potential importance of these channels, this work proposes an additional perspective and attempts to uncover the implications of own experience of immigration, encountered acceptance, and assigned status in the host society on the current immigration attitudes of established immigrants.

To investigate this idea, I use a survey-experiment with a sample of participants with immigration background residing in Germany and experimentally vary the status of the participants' in-group. In a separate pre-study, a smaller group of participants from the majority population, that is, those with no immigration background, is asked to evaluate different immigrant groups (structured along the region of their origin) as contributing rather positively or rather negatively to "the socio-economic and cultural life in Germany". In the second and main part of the experiment, a sample of participants with immigration background ($n = 1,159$) is presented a subset of answers elicited in the first phase. Participants are randomly chosen to be presented a subset of answers that evaluates their in-group either positively or negatively, while holding the evaluation of two other

out-groups constant. The statement was designed to deliver prejudiced evaluations of the three groups and manipulate the status position of participants in-group in the fixed ethnic hierarchy. I investigate the effect of the randomly assigned evaluation of the own in-group on the thereafter expressed support for refugees from the Middle East, captured by the respondents' willingness to forgo some part of their experimental earnings in order to secure a donation to the United Nations High Commissioner for Refugees (UNHCR). Additionally, several attitudinal measures of participants' position towards refugees are also elicited.

The tendency of individuals to classify themselves and others into in- and out-groups, as well as the competition for status is a well-documented and seemingly universal characteristic of human societies. Social Identity Theory (SIT) (Tajfel et al. (1979)), starting from the assumption that individuals strive to enhance their self-esteem, offers an explanation of inter-group dynamics in the presence of a group-based identity threat. According to SIT, the lower the status assigned to a group, less can it contribute positively to its members' social identity. In order to cope with the identity threat, the members of such a group are predicted to engage in defensive strategies, whereby the contextual factors stand to determine which of the defensive strategies are available in any given case. If group boundaries are sufficiently permeable, a member of a low-status group might attempt to disassociate from the group and join a more favorably evaluated group. If the appearance of a new immigrant group recasts the established immigrants group(s) as less distant from the native majority than what was previously perceived, and thus softens the boundary between these two groups, it might open a way for this strategy, which Tajfel and Turner termed as "individual mobility". However, if the group boundaries are firm enough to not allow for individual mobility between the groups, members of the low-status groups might attempt to seek relief in changing the out-group relevant for the comparison. In particular, members of a low-status group are predicted to recover self-esteem by focusing on comparison with an even lower status group and by emphasizing the own group's positive distinction relative to this group. Thus, in the context observed here, both proposed strategies of coping with identity threat predict a disassociation with an overarching category of

immigrants on the part of established immigrants, and potentially even an outright rivalry with new-coming lower-status immigrants.

The mechanism studied in this work particularly resemble the so-called Queen-Bee phenomenon. The term, as described in Ellemers et al. (2004), should designate women occupying positions in male-dominated environments, who express a gender bias against women in evaluating their lower ranked female subordinates, sometimes even more so than their male colleagues, while at the same time distancing themselves from their own gender by expressing masculine self-descriptions. Subsequent work in this literature (for review, see, e.g. Derks et al. (2016)) has relied on both social identity theory and the system justification theory (Jost (2019), Jost et al. (2003)) to argue that rather than being a behavioral trait specific to women, the Queen-Bee behavior is in itself a response to the gender bias and identity threat in the male dominated environments. Drawing a parallel with the question considered here, one might wonder if there exists a Queen-Bee-Immigrant phenomenon. That is, do the established immigrants respond to an environment sceptical toward immigrants by distancing themselves from the immigrant status and expressing negative bias toward new immigrants. While the Queen-Bee literature considers a bias of females toward other females, that is toward own in-group, reacting by being more suspicious of the other (immigrant) out-group should arguably be even less psychologically costly, and thus more likely strategy.

The experimental results support these predictions. Participants who received a negative evaluation of the own in-group donated systematically less to the UNHCR, compared to the participants who received a positive evaluation. The difference in donations amounted to 4.4 euros, representing around 12% of the average donation. This result is not explained by the demographic characteristics, region of residence (in Germany), or region of origin of established immigrants. Additionally, a collected measure of participants' post-treatment mood shows that the treatment effect is not propagated through its effect on participants' mood.

In the next step, I study the treatment effect on participants' perceived descriptive norm surrounding the expression of prejudice in the majority population, and investigate whether the effect on norms mediates the effect on willingness to donate. Previous works

on the emergence of social norms showed that individuals, at least in part, infer the group's descriptive norms (what others are doing) from other individuals' behavior to which they are incidentally exposed. In particular, in situations where the behavior of interest does not produce an easily observable outcome (such as litter in public space), people combine summaries of group's behavior (e.g., election outcomes), with the direct experiences that they make to learn the descriptive norm regarding this behavior (Kwan et al. (2015), Kashima et al. (2013)). Thus, it could be hypothesized that the groups that were socialized in the presence of a steep ethnic hierarchy and were exposed to prejudiced treatment in the course of their integration grow to perceive inter-ethnic competition and expression of prejudice as pervasive, and perhaps even legitimate social dynamics in the host society, and are more likely to apply it towards the lower ranked groups once they encounter them. In particular, I hypothesize that exposing established immigrants to a lower acceptance, that is to a negative evaluation of the own in-group, expressed by a (high-status) majority member, updates their perceived descriptive norm such that they perceive expressing lower acceptance towards low-status groups as more frequent among the native majority.

In order to test this prediction, I elicit participants' empirical beliefs regarding the percentage of the pre-study participants who negatively evaluated the impact of several different immigrant groups, including refugees from the Middle East, one (in Germany) salient and one non-salient low-status immigrant out-group (immigrants from Turkey, and those from south of Africa), as well as one high-status out-group (immigrants from western European countries). Experimental results provide support for the prediction. Participants exposed to a lower acceptance, that is, those who received negative evaluation of the own in-group expected systematically more negative evaluations of all low-status out-groups (but not of the high-status one) on the part of the majority population participants. Whereas the treatment effect on injunctive norms (what others believe one oath to do) is not explicitly tested here, the literature on social norms provides ample evidence for the role that descriptive norms alone play in shaping intentions and behaviors (Bicchieri and Xiao (2009), Krupka and Weber (2009),

Bardsley and Sausgruber (2005)) in a wide range of behavioral domains, including expression of prejudice (Álvarez-Benjumea and Winter (2020)).

Previous works studying how privately held opinions translate into publicly expressed attitudes and behaviors found that stigmatization and social desirability of certain beliefs play an important role in determining to which degree the discrepancy between the two emerges. In particular, individuals tend to bias their statements when publicly expressed towards positions deemed socially more appropriate (Bursztyn et al. (2018), Perez-Truglia and Cruces (2017), Enikolopov et al. (2020)), or those that are more typical of the group with which they identify (Janus (2010)). As opinions on immigration and asylum represent a typical example of the sensitive attitude that might be prone to misrepresentation, it is interesting to study whether established immigrants express different attitudes towards refugees privately than when these attitudes are potentially observable.

The intuition underlying misrepresenting attitudes is individuals' desire to present themselves in a way that they believe would be appreciated by those observing. The direction of misrepresentation is therefore informed by the differences between own and expected others' positions. However, in a situation where the majority is polarized over questions of immigration, and there are no clear expectations of immigration attitudes of most established immigrants groups, it is not clear in which direction the latter would feel tempted to skew their opinions if observed by the former. The works of Fouka et al. (2021) and Fouka and Tabellini (2021) illustrate how in a context of a country populated by a high-status majority and some immigrant minority groups, an appearance of a new immigrant group can trigger the redistribution of status and open a possibility for the groups to be re-positioned in the hierarchy. Particularly, they show on two separate examples how an influx of a new and salient immigration group in the U.S. (African Americans during the Great Migration and Mexicans in the nineties) led native whites to perceive a lower social distance to established immigrants and re-categorize them as in-group. What remains unanswered is whether the established minorities also perceive the appearance of a new group as a possibility for improving their own status and respond to it by signalling their concern regarding the newcomers

to the majority population, specifically if they perceive the majority population to be skeptical towards the newcomers.

In order to investigate these questions, participants are asked to provide an answer to one attitudinal question regarding refugees in Germany once privately, and once where participants are aware that a future participant, randomly selected from a sample of majority population participants, might observe their answer. Comparing the answers provided in both settings reveals that participants indeed do answer differently when potentially observed, and the direction of misrepresentation depends largely on the initial, privately expressed preference. In particular, participants who provided a more critical assessment of the impact of refugees in Germany when answering privately changed their answer towards expressing more supportive views in the observable setting. More interestingly, the opposite holds for the participants who privately assessed the impact of refugees more positively, that is, they misrepresent their positions in the observable setting so as to appear more critical. Furthermore, among participants who were more critical in the private setting, those assigned to the Negative treatment misrepresent their attitudes in the observable setting (in the positive direction) systematically less than those in the Positive treatment.

Finally, I explore the role of the reciprocity preferences as a potential channel for the effect that providing different evaluations of the immigrants' in-group has on their support for the refugees. In particular, upstream indirect reciprocity designates a tendency of individuals to exhibit prosocial behaviour towards others because somebody else has exhibited prosocial behaviour towards them (Alexander (1987), Nowak and Sigmund (2005)). I elicit participants' preferences for upstream indirect reciprocity in an extended dictator game and provide evidence for its effect in line with the theoretical prediction. Participants with a higher preference for reciprocity donated more and were more likely to make a positive donation if they were in positive treatment, and donated less (though insignificantly) and were less likely to make a positive donation if they were in the negative treatment.

This work contributes to the literature on the political preferences of immigrants (Dinas et al. (2021a), Strijbis and Polavieja (2018), Van der Zwan et al. (2017), Just

and Anderson (2015), Dancygier and Saunders (2006)), and more specifically to the branch studying how political attitudes of the native majority shape these preferences (Dinas et al. (2021b), Fouka (2019), Kuo et al. (2017)). To the best of my knowledge, this is the first paper that provides causal evidence for the effect of status deprivation (through expressed prejudice) on immigration attitudes of the immigrant population. More generally, this work contributes to the broad literature on immigration attitudes and the drivers behind them (for survey, see Hainmueller and Hopkins (2014)). Finally, this paper also relates to the discussion on political correctness, by highlighting the negative externalities entailed by its absence in the inter-ethnical context (Braghieri (2021), Norton et al. (2006), Morris (2001)).

2.2 Experimental Design

The study is split into two phases, which will henceforth be referred to as the pre-study and the main experiment, both implemented as an online survey. In the following, I provide the description of both phases.

2.2.1 Pre-study

The pre-study was conducted with a small sample ($N = 125$) of participants residing in Germany and with no immigration background. The only purpose of the pre-study was to collect the responses from the majority population regarding their evaluations of different immigration groups that would later be used in the main experiment.

At the beginning of the survey, participants provided answers to a set of basic demographic questions, including participant's gender and age, alongside own and parental country (countries) of birth, which were used to ensure that only participants from the majority population with no immigration background, participate in the pre-study.

Thereafter, for each of the several regions/countries, participants were asked to evaluate whether people immigrating from the given region/country contribute rather positively or rather negatively to socio-economic and cultural life in Germany (participants selected one of the two options as an answer). To avoid confusion in terms of which countries are encompassed by a given region, with each question participants were shown a simple political map of the relevant part of the world, with the region of interest visibly highlighted, and with the text of the question explicitly listed all corresponding countries. An example question and the exact phrasing of the questions is available in the Appendix. Participants in the pre-study were paid a fixed participation fee upon completion of the survey.

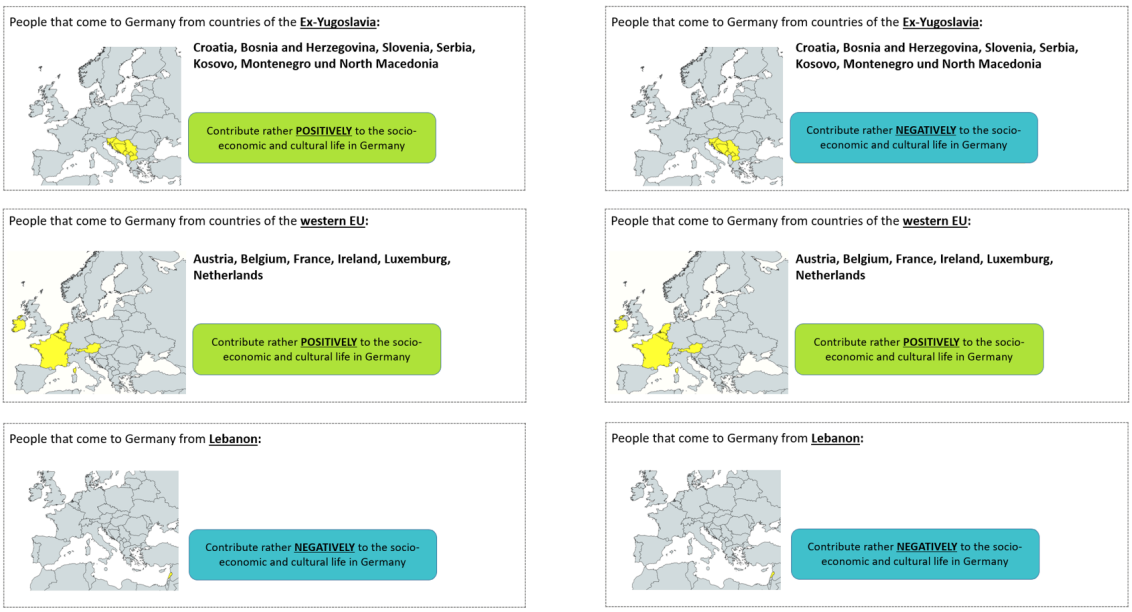
2.2.2 Main experiment

The main part of the experiment was conducted with a sample of 1.175 participants with immigration background residing in Germany.

Demographics As in the pre-study, at the beginning of the session, participants answered the questions regarding their basic demographic characteristics, including participants' own and parental country of birth. This information was used to match participants to one of the eleven regions or origin.[1]

Treatment provision In this part of the experiment, participants are told that, in a study that took place at an earlier point of time, a group of 125 participants from Germany with no immigration background were asked to evaluate the impact of various immigrant groups on socio-economic and cultural life in Germany, and that some of the collected answers will be shown to them. Participants are then (conditional on the region that they were matched to) randomly split into two treatments. Participants in both treatments are presented with one evaluation of each of the three immigrant groups - one representing immigrants stemming from their own (parental) region of origin, and the other two representing two out-groups. In both treatments, the answers from the pre-study are selected so that one out-group (in both treatments: immigrants from western EU countries) is always evaluated positively and the other (in both treatments: immigrants from Lebanon) negatively. Here, the positive and negative evaluations refer to the group being evaluated as "contributing rather positively", and respectively as "contributing rather negatively", to the socio-economic and cultural life in Germany. The only difference between the treatments is the evaluation of the own in-group. In the **Positive treatment**, participants are shown an answer that evaluates the impact of the own in-group positively, whereas in the **Negative treatment**, participants are shown an answer that negatively evaluates the impact of the own in-group. Figure 2.1, provides an example of evaluations presented to participants for both Positive and Negative treatment.

[1]The eligible regions of origin in this study included: Countries in central-eastern European Union (Czech Republic, Slovakia, Poland, Hungary); Romania and Bulgaria; Baltic states (Estonia, Latvia, Lithuania); Countries of ex-Yugoslavia (Bosnia and Herzegovina, Croatia, Kosovo, Montenegro, North Macedonia, Serbia, Slovenia); North Africa (Morocco, Algeria, Libya, Tunisia, and Egypt); Southern European Union countries (Greece, Italy, Portugal, Spain, Cyprus, and Malta); Turkey; Southern countries of the ex-Soviet Union (Tajikistan, Turkmenistan, Georgia, Kazakhstan, Kyrgyzstan, Armenia, and Azerbaijan); Western countries of the ex-Soviet Union (Ukraine, Moldova, Belarus); Russia; and Albania. The division was made with the aim of including the regions of origin most frequently encountered among the population with immigration background in Germany. At the same time, the division attempted to achieve a trade-off between the number of regions and a sufficiently narrow definition of a region so as to allow for successful clustering.



(a) Positive treatment (b) Negative treatment

Figure 2.1. Treatment provision - example

The figure depicts an example of a screen that a participant, who was matched to the region of Ex-Yugoslavia, would see in the treatment provision phase if they were allocated to the Positive treatment (panel), and that if they were allocated to the Negative treatment (panel b). Participants are informed that they would see a subset of answers collected in the pre-study. Treatment variation is based on randomly matching participants to an answer from the pre-study evaluating participant's own (parental) region of origin either positively or negatively while keeping the evaluations of the other two out-groups constant.

Including the two out-groups, consistently evaluated positively and negatively, ensures that provided information cannot be interpreted as a more or less positive attitude towards immigration in general, and instead ties the treatment variation to the position of the own in-group in an already set hierarchy.

Elicitation of attitudes towards refugees In this part of the study, two measures of participants' support of refugees were elicited. Following the approach of Dinas et al. (2021a), an attitudinal measure of support was constructed by collecting participants' answers to a set of seven questions. Participants provided their views (among others) on whether Germany should increase or decrease the number of people it grants asylum to, refugees' influence on the labor market, the welfare state, probability of a terrorist attack, criminality, etc. The list of all questions is provided in the Appendix.

The main behavioral measure of participants' support for refugees was captured by the willingness to donate to the United Nations High Commissioner for Refugees (UNHCR).

Participants were informed that, as a part of the study, a lottery would be administered whereby one randomly selected participant will be awarded 100 euros and all participants have the same chance of winning the prize. They are then asked whether they would like to donate some part of the 100 euros prize, in the case that they win the lottery, to the UNHCR, and if so, how much. Participants are informed that if they decide to dedicate some amount to refugees-support, this amount will be automatically deducted from their 100 euro prize in the case they win, and a donation in the same value will be made to an organization supporting refugees.

Mood elicitation In order to be able to control for the treatments' potential effect on participants' mood, a measure of mood is elicited via Self-Assessment Manikin questionnaire (Bradley and Lang (1994)). Three questions, intended to capture three major affective dimensions - pleasure, arousal, and dominance - asked participants to select one of the five offered manikins that they felt best describes their mood.

Empirical expectations In order to study treatment effects on participants perceived descriptive norms regarding prejudice expression, in this part of the experiment, participants were asked to guess what percentage of the 125 participants without immigration background that took part in the pre-study evaluated negatively each of several immigrant groups (categorized by their region/country of origin). Each participant was asked to guess the share of participants from the pre-study who negatively evaluated the impact of people immigrating to Germany from: participant's own (parental) region of origin, western countries of the European Union (Austria, Belgium, France, Ireland, Luxemburg, Netherlands), Lebanon, Turkey, countries of southern Africa (South African Republic, Namibia, Eswatini and Lesotho) and that of refugees immigrating from the Middle East (Syria, Iraq, Afghanistan, and Pakistan). Countries within a given region were visibly displayed to participants. Participants were informed that the answer closest to the true collected values would be rewarded by additional 25 euros.

Indirect upstream reciprocity One potential driver of treatment effects on expressed support for refugees might be the preference for upstream indirect reciprocity, that is, the tendency of individuals to exhibit prosocial (antisocial) behaviour towards

others because somebody else has exhibited prosocial (antisocial) behaviour towards them. To facilitate studying this mechanism, in this part of the experiment, a measure of indirect upstream reciprocity was collected using an extended dictator game with three players. Each Participant is assigned one of the three roles: player A, player B, or player C. Thereby, player A is given a budget of 30 euros, out of which they can send a certain sum to another player B, who in turn can send some of the received amount to player C. The amount sent by player A is multiplied by a factor f, and the resulting amount is paid to player B. Player A and player B known that the multiplication factor can take either a high value ($f = 4$) or a low value ($f = 2$), but the realization of this value is not know to any of the players. Thus, player B observes only the resulting sum they received but is not aware whether it resulted from player A sending a higher sum that was multiplied by a low factor value, or from player A sending a lower sum that was multiplied by a high factor value. Here, player A could select between sending 0, 8, 16, and all 30 euros. All participants assigned to role B received a total of 32 euros (corresponding to player A sending either 8 or 16 euros, and the factor being equal to either 4 or 2, respectively).

Player B is then asked to decide for both scenarios how much of the received sum they would like to send to person C. To ensure that welfare concerns do not play a role in the decision of player B, the amount sent to player C is paid to them without multiplication. Participants are informed that at the end of the study, one triplet will be selected and paid out the amounts according to the decisions they made. Most of the participants were assigned the role of player B ($n = 1164$), and the rest was distributed among the other two roles.

I take the difference in amount sent to player C in scenario where player A was more generous versus that when they were less generous as a measure of indirect upstream reciprocity of player B.

Preference falsification When individuals' are asked to state their political views while observed by the others, preference falsification might mask truly held preferences and skew them to the perceived socially appropriate positions. This part of the experiment has the aim to capture a potential difference in attitudes expressed by

established immigrants when they expect these attitudes to be observed by a majority population, as compared to when this is not the case.

In this part, participants are reminded that all previously provided answers will be delivered only to the researchers in anonymized form. The participants are then informed that only in this part of the experiment they are asked to provide an answer that can be used in a potential future study to inform future participants about their views on immigration. Furthermore, the instruction clarifies that, if the future study is conducted, it will be run in Germany with a sample of German citizens and that the recipient of their answer would know their country (countries) of origin. Thereafter participants fill out the answer to the question "Is Germany made a worse or a better place to live by refugees who are granted asylum in Germany?", that was already asked as one of the attitudinal questions in the "Elicitation of attitudes" phase.

Preference for equal treatment Should established immigrants feel that the refugees are exposed to lighter immigration policies requirements compared to those in their own immigration experience, this might lead to a perception of unfair treatment and reduce their support for refugees. This part of the experiment attempted to elicit participants' preference for equal treatment, that is, whether participants, after being exposed to certain conditions, would approve if the conditions were improved for other participants. Participants are asked to provide an answer to one cognitive reflection question (Frederick (2005)) and are informed that the participant who provides a correct answer in shortest time will be rewarded by additional 30 euros. After providing the answer (but before learning the outcome), participants are told that the same question might be used again in a (potentially implemented) future study to measure the numerical ability of participants. They are then asked to vote on how should the question, if used, be incentivized. Participants choose between keeping the reward as in the current experiment (fastest correct answer rewarded with 30 euros), rewarding additionally the second-fastest correct answer by the same sum, rewarding the first three correct answers by the same sum each, or rewarding only the fastest correct answer by a decreased sum of 20 euros. Participants are told that if the question is used in the future study, the option that received the most votes will be implemented.

Additional demographics and debriefing At the end of the experiment, participants are shown the true percentages of participants in the pre-study who negatively evaluated each of the several groups. The session ended after collecting some additional basic demographic information.

2.2.3 Data and sample description

The study was conducted in the period December 2021 to January 2022. The sample for the pre-study involved 125 adult individuals with residence in Germany and with no immigration background. A participant was considered to have an immigration background if they, or at least one of their parents, was born outside of Germany. For the purposes of the main-experiment, a separate sample was recruited involving 1,175 adult individuals with residence in Germany and with an immigration background. Out of this number, 16 participants provided inconsistent answers to basic demographic questions (stated unreasonable age), and their answers were removed, resulting in a sample of 1,159 participants.

Participants with an immigration background were matched to what I will be for simplicity referring to as "region of origin", indicating one of the eleven regions encompassing their, or parental, country of birth. The regions selected to be targeted in this study encompassed all countries within Europe (except for the Western European countries), all Ex-Soviet countries, Turkey, and five northern African countries (Egypt, Tunis, Morocco, Algeria, and Libya). Table 2.7 in the Appendix provides an overview of all regions (and all encompassed countries), along with the share of participants matched to each region. The selection of the eligible regions attempted to match the studied sample with the groups most represented among the population with an immigration background in Germany[2], and to focus on those immigrant groups that are more likely to occupy a lower status position in German society (thus the exclusion of the Western European countries). Table 2.8 in the Appendix presents the descriptive statistics of the sample across both treatments. The online survey was programmed in

[2]See Statistical Office of Germany (Genesis-Online Database, code: 12211-0202)

Qualtrics and the distribution of the link to the experiment was delegated to a panel company CINT.[3]

In the next section, I provide the overview of empirical results and test the following (pre-registered) hypotheses:

Hypothesis 1: Being assigned to the Negative treatment leads to a decrease in the amount donated to UNHCR and a more negative evaluation of refugees as measured by the attitudinal questions.

Hypothesis 2: Being assigned to the Negative treatment leads participants to expect a higher percentage of negative evaluations of refugees' impact on socio-economic and cultural life in Germany among majority participants (in the pre-study). Furthermore, assignment to the Negative status treatment leads participants to expect a higher percentage of negative evaluation of the own in-group, as well as of the other low-status groups among majority participants.

Hypothesis 3: Participants with higher indirect reciprocity react more strongly to treatment variation, that is, express more negative (positive) evaluations of refugees in the Negative (Positive) treatment.

Hypothesis 4: The distribution of answers provided to the question "*Do refugees who obtain asylum right in Germany make Germany a worse or a better place to live*" in "private" scenario differs from the distribution of answers provided to the same question in "observable" scenario. Furthermore, being assigned to the Negative treatment leads participants to express a less favorable opinion of refugees in "observable" scenario.

[3]https://www.cint.com/

2.3 Results

2.3.1 Pledged donation to the UNHCR

In this subsection, I present the measured effect of the treatment, that is, the effect of receiving negative status information, compared to receiving positive status information, on both behavioral and attitudinal measures of support for refugees. The behavioral measure of support for refugees is captured by the amount that participants committed to donate to the United Nations High Commissioner for Refugees (UNHCR), from a 100-euro prize that is raffled among all participants at the end of the study. On average, participants committed to donate 36.86 euros, with individual decisions spanning across the full range of possible donations. Figure 2.2 provides an overview of the observed distribution of pledged donations.

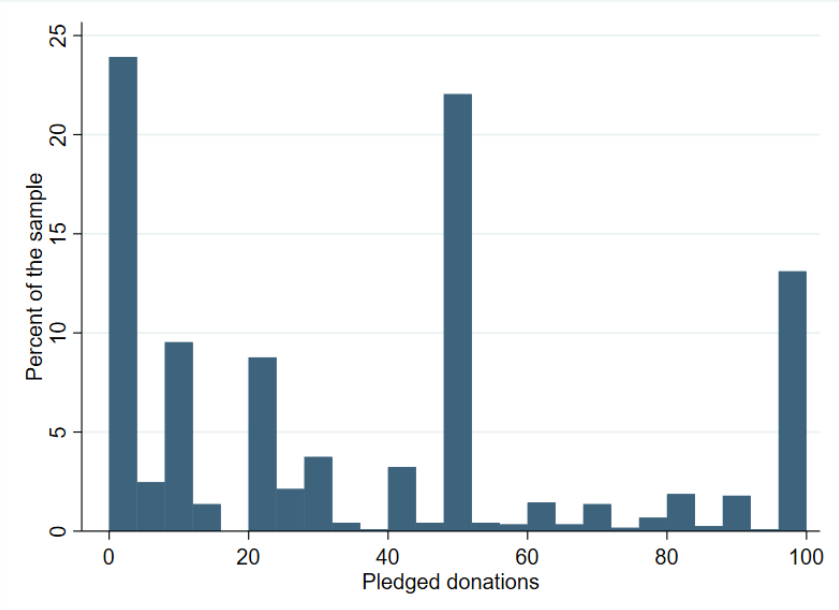


Figure 2.2. Distribution of pledged donations to the UNHCR

The results presented in Table 2.1 depict the effect of being allocated to the Negative treatment (with Positive treatment serving as a baseline) on the pledged donations. Considering that the possible value of the donation was limited at 0 from below, and at 100 from above, and that the number of participants who selected both limiting values was significant, the table presents the results of Tobit regression of the donated amount on treatment variable and individual controls. All presented regressions include fixed

effects of the federal state within Germany and region of origin, and the standard errors are clustered on the level of participants' region of origin.

Table 2.1. Treatment effects: Pledged donation to the UNHCR

	(1)	(2)	(3)	(4)
Negative treatment	-7.049***	-6.922***	-0.189***	-0.190***
	(1.593)	(1.532)	(0.063)	(0.060)
Constant	47.418***	54.824***	1.031***	1.400***
	(3.082)	(4.608)	(0.152)	(0.212)
Marginal effects: $E(\Delta y/\Delta x)$				
Negative treatment	-4.716***	-4.630***	-0.054**	-0.054**
	(0.000)	(0.000)	(0.003)	(0.001)
Individual controls	No	Yes	No	Yes
Observations	1,159	1,159	1,159	1,159

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.
Column (1) and column (2) show Tobit regression of amount dedicated to donate to the UNHCR on treatment variable and the set of individual controls. Negative treatment indicates receiving negative status information regarding own in-group (with Positive treatment serving as a baseline). Reported marginal effects represent the average marginal effect of being allocated to Negative treatment on donated amount. Columns (3) and (4) show Probit regression of of an indicator variable for donation being larger than zero on treatment variable and the set of individual controls. All regressions include fixed effects of the federal state of residence in Germany and region of participants' (parental) origin. Individual controls (included in columns (2) and (4)) include age, gender, equivalent household income tertile and indication of tertiary education. Reported marginal effects represent the average marginal effect of being allocated to Negative treatment on probability of making a positive donation, and can be directly interpreted in terms of percentage points difference. Standard errors in parentheses are clustered on the level of region of participants' (parental) origin.

The results in Table 2.1 provide support for the Hypothesis 1. The results shown in column (1) demonstrate that participants in the Negative treatment commited to donate systematically less to the UNHCR. Participants pledged on average around 4.5 euros less to donation if they were in the Negative treatment ($p < 0.01$), which represents around 12% of the average committed sum. Furthermore, as shown in column (3), participants allocated to the Negative treatment were significantly less likely to pledge any positive donation relative to those in the Positive treatment. In particular, reallocating a participant from Positive to Negative treatment decreased, on average, the probability of the participant pledging a positive donation by 4.9 percentage points ($p = 0.004$). The results in columns (2) and (4) show that these

findings are robust to the inclusion of controls for the respondents' socio-demographic background.

The collected measure of participants' mood allows checking whether the treatment variation affected the behavior through its effect on participants' mood. However, the distribution of all three measured affective dimensions - pleasure, arousal, and dominance, elicited using the Self-Assessment Manikin questionnaire (Bradley and Lang (1994)), did not differ significantly between the two treatments (Kolmogorov–Smirnov test for equality of distribution in both treatments, for each of the three affective dimensions - pleasure: $p > 0.6$; arousal: $p > 0.9$; dominance: $p > 0.9$). Furthermore, Table 2.9 in the Appendix shows that, while valence and arousal had a positive effect on willingness to donate, controlling for these measures does not qualitatively alter the observed effect of the treatment.

2.3.2 Attitudinal measures

In addition to the behavioral measure of support for refugees, a set of attitudinal measures was elicited by means of collecting answers to seven questions regarding refugees from Syria, Afghanistan, Iraq, and Pakistan who flee to Germany. The questions, among others, regraded participants' views of the influence of refugees on employment, risk of terrorism, criminality. The exact formulation of all seven questions is provided in the appendix.

Compared to treatment effect on pledged donations, treatment had a smaller effect on the attitudes reported in the seven questions. The first six columns of Table 2.2 show the results of ordered logistic regression of chosen answer for each of the (first six) questions on treatment variable and the set of socio-demographic controls. All answers are re-coded such that a higher value indicates higher support for refugees. Column (1) shows that in the case of the first question, which asked the participants' opinion on whether Germany should increase or decrease the number of people it grants the asylum to, participants were significantly more likely to provide a lower answer (decrease number of granted asylums) if they were in the Negative treatment. However, although treatment effects

Table 2.2. Treatment effects: Attitudinal questions

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	q1	q2	q3	q4	q5	q6	q7	$\bar{q}$
Negative treatment	-0.266**	-0.082	-0.051	-0.031	0.043	-0.144	-0.049	-0.107
	(0.109)	(0.105)	(0.122)	(0.111)	(0.126)	(0.089)	(0.040)	(0.105)
Constant							1.082***	2.120***
							(0.174)	(0.178)
Individual controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1,159	1,159	1,159	1,159	1,159	1,159	1,159	1,149

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.
Columns (1) through (6) show the results of ordered logistic regression of provided answer on treatment variable and the set of individual controls. Column (7) shows the result of Probit regression of dummy variable that takes value 1 if a participant selected "To flee war" or "Avoid political persecution" as primary reason why refugees leave their countries, and 0 otherwise. All regressions include fixed effects of the federal state of residence in Germany and region of participants' (parental) origin. Individual controls include age, gender, equivalent household income tertile and indication of tertiary education. Standard errors in parentheses are clustered on the level of region of participants' (parental) origin.

work in the predicted direction in most of the other questions (that is, participants in the Negative treatment provided less supportive answers), these effects are not significant.

Question q7 asked participants to provide their opinion on the primary reason why refugees abandon their countries among the following options: "To flee war", "Avoid political persecution", "Improve their economic conditions" and "Obtain access to social security payments in the destination country". I construct a dummy variable that takes value one if a participant selected one of the first two choices and show in column (7) the results of Probit regression of this variable. Again here, being assigned to the negative treatment decreased the probability of selecting one of the two reasons that would indicate security (rather than economic) concerns as a primary reason for flight, but the effect is insignificant.

Finally, for each participant, I construct an aggregate measure of answers to attitudinal questions by averaging seven dummy variables. The dummy variables correspond to the seven questions, and each takes value one if the participant selected an answer to the respective question that indicates higher support for refugees than that implied by the neutral point (selected 3 (5) on a scale 1 to 5 (0 to 10)). Column (8) shows the results of regressing this aggregate measure, denoted by $\bar{q}$, on the treatment variable and the set of individual controls.

2.3.3 Empirical expectations - differential evaluation based on origin

Results in the previous section showed that providing participants with a negative evaluative opinion on immigrants from their own (parental) region of origin, expressed by a member of the majority population, led them to significantly decrease their support for refugees. One possible explanation for this regularity might be that participants, who face differential acceptance by the majority population based on their origin, might internalize this behavior as usual and perhaps legitimate in the society more generally. In other words, people from low-status regions could learn from discrimination directed towards their own in-group that discriminating downwards (i.e., against groups ranked lower than one's own group) is widespread and possibly also acceptable behavior in the host society. As proposed by Hypothesis 2, in the context observed here, this would suggest that observing lower acceptance of the own (lower-status) in-group might negatively update participants' empirical expectations of acceptance of other lower-status groups, such as refugees, among the majority population.

In order to test this prediction, I collected an incentivized measure of empirical expectations on approval of different immigrant groups by the majority population. After collecting the main outcomes of interest, participants were asked to guess the share of respondents in the pre-study (without migration background) who evaluated *negatively* the impact of each of several immigrant groups on socio-economic and cultural life in Germany. Particularly, each participant was asked to guess the share of participants from the pre-study who negatively evaluated the impact of people immigrating to Germany from: participant's own (parental) region of origin, western countries of European Union, Lebanon, Turkey, countries of southern Africa and that of refugees immigrating from the Middle East. To avoid confusion, in cases where the evaluation regarded people immigrating from a given region, all countries within the region were listed. The exact phrasing of the question and an example screen seen by participants is provided in the Appendix.

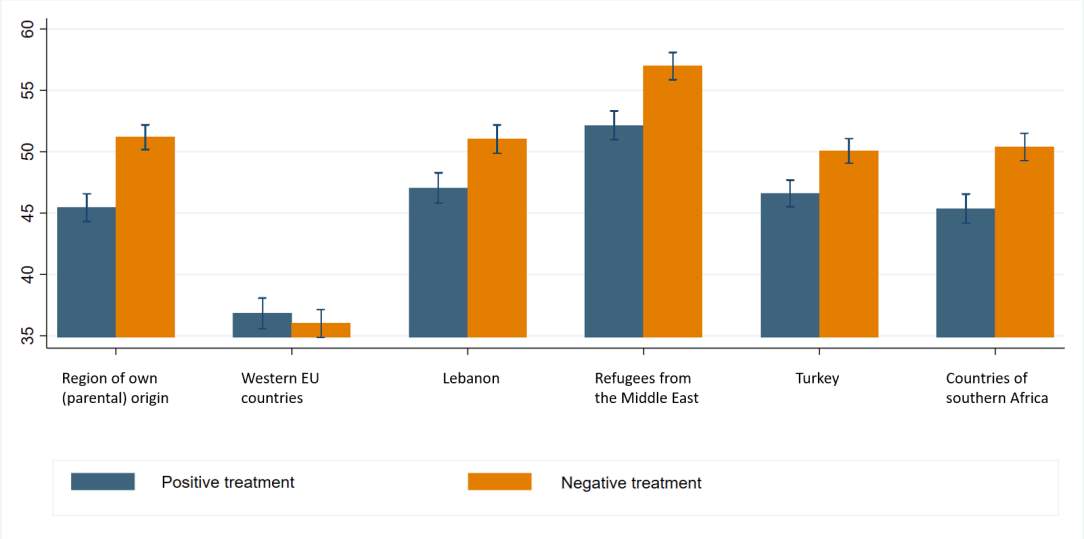


Figure 2.3. Treatment effect on perceived descriptive norm

The figure depicts the average elicited guesses of the share of majority population participants who evaluated negatively the influence on socio-economic and cultural life of people immigrating to Germany from countries/regions depicted on x-axis, by treatment. The vertical lines indicate the 95% confidence intervals.

Figure 2.3 provides an overview of measured treatment effects on collected empirical expectations. The first pair of bars on the left shows that participants who received a negative evaluation on their own in-group, on average, expected the majority population participants to be more critical towards immigrants from their region of origin. This is also intuitive, as it reflects the information that participants received in treatment provision, but is still informative as it shows that participants extrapolated from the individual evaluation that they received to the average opinion of the group. At the same time, it serves to confirm the successful treatment manipulation.

More interestingly, the same applies to participants' expectations of evaluations of all other low-status immigrant groups. Particularly, in accordance with Hypothesis 2, participants in the Negative treatment expected a significantly more negative evaluation of the impact of refugees from the Middle East, as well as of people immigrating from Turkey, Lebanon and from countries in the south of Africa. This is not the case for the expected evaluation of high-status immigrants, that is, those coming to Germany from the western EU countries, indicating that this is not a consequence of expecting the majority population to be more skeptical towards immigrants in general. Instead, as proposed by Hypothesis 2, it appears that receiving a negative evaluation of the own

in-group led participants to expect more critical views only of those immigrant groups that were of a lower status than those who are evaluating.

Table 2.3. Treatment effects: Empirical expectations

Elicited expectation:	What percentage of majority population participants evaluated negatively the impact of people coming to Germany from:					
	(1)	(2)	(3)	(4)	(5)	(6)
	Own (parental) region of origin	Refugees	Turkey	Lebanon	Southern Africa	Western EU countries
Negative treatment	5.863***	4.921**	3.524**	4.184*	5.112*	-1.463
	(1.118)	(2.187)	(1.161)	(2.021)	(2.444)	(1.330)
Constant	50.237***	53.229***	50.914***	38.812***	37.494***	41.056***
	(3.080)	(4.129)	(4.560)	(4.277)	(3.734)	(2.375)
Negative treatment	5.763***	4.882**	3.708***	4.023*	5.187*	-1.252
	(1.034)	(2.101)	(1.135)	(1.925)	(2.401)	(1.393)
Individual controls	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1,159	1,159	897	1,159	1,159	1,159

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Robust standard errors in parentheses.
OLS regression of the elicited guess (of participants with migration background) of the share of majority population participants who evaluated as negative the imact on socio-economic and cultural life in Germany of people immigrating to Germany from countries/regions shown in columns' headers. First column regards the people who immigrate to Germany from country/region of participant's origin (or that of their parent(s) if the participant was born in Germany). In questions that regarded immigrants from a region (rather than a country) all countries within the region were listed in the question. All regressions include fixed effects of the federal state of residence in Germany and region of participants' (parental) origin. Standard errors are clustered on the level of region of participants' (parental) origin. Individual controls include age, gender, equivalent household income tertile and indication of tertiary education.

I investigate these observations more formally in Table 2.3, which provides the results of OLS regression of the participants' estimates of the share of the majority population participants who negatively evaluated each of the mentioned immigrant groups. Confirming the indications provided by Figure 2.3, regression results show that participants in the Negative treatment (relative to those in the Positive treatment) expected significantly more negative evaluations of the impact of refugees, as well as of all low-status immigrant groups, but not of the high-status one. The results are significant and are not explained by participants' socio-demographic characteristics.

Whereas the main outcome of interest here was a spillover effect on the participants' empirical expectations regarding refugees from the Middle East, it is particularly

interesting to note that the spillover affected not only a salient unrelated minority (Turkish), but also a very non-salient group of immigrants from countries in the south of Africa, who are effectively barely present in Germany, both as a share of population[4] and in the public discourse. The question regarding immigrants from this region explicitly specified the countries in question (South Africa, Namibia, Eswatini, and Lesotho), thus this effect can not be the consequence of mistaking this region for other regions/countries in Africa where some percentage of the refugees came from (e.g. Eritrea). On the other hand, considering that some refugees indeed did come from some Sub-Saharan countries, it might be that the spillover effect from own in-group evaluation to the evaluation of the refugees, further spilled-over to any group that remotely resembles this group, even if the only commonality between the groups is the same continent of origin. It, therefore, illustrates how social dynamics, completely unrelated with the characteristics of the immigrant group in question, can pre-set the stage and shape attitudes towards this group, even before it is present in the host country.

While the literature on social norms provides evidence of an impact of empirical expectations (what others are doing) on normative expectations (what others believe one ought to do) (see, e.g., Bicchieri et al. (2020)), the results provided here can only support the treatment effect on the former. Therefore, whether the experience of being deferentially evaluated based on the place of origin shapes as well the perceived appropriateness of such behavior remains an interesting open question.

On the other hand, irrespective of their influence on normative expectations, empirical expectations have been shown to influence behavior in a wide range of domains (Bicchieri and Xiao (2009), Krupka and Weber (2009), Bardsley and Sausgruber (2005)). In order to test whether the effect of receiving a negative evaluation on the own region of origin on donations was mediated by its effect on empirical expectations regarding the evaluation of refugees, I instrument the expectations by treatment variable and run a two-stage least square regression. Table 2.4 shows the results of the second-stage regressions, both

[4]According to the data of Federal Statistical Office of Germany (Genesis-Online Database, code: 12521-0002), at the end of 2020, the number of people residing in Germany with citizenship of one of these four countries is below 8500 persons.

for the donated amount and for the probability of making a positive donation. The depicted results suggest that, assuming the absence of factors that could jointly cause the expectations and the donation, the treatment effect on donation indeed worked through its impact on expectations. Whereas the ultimate test for the behavioral effects of the empirical expectations would amount to administering a norm-manipulation experiment and is thus outside the scope of this work, these results can be taken as tentative evidence for the mediating role of expectations.

Table 2.4. Mediating effect of empirical expectations

	(1)	(2)
	Donation	Pr(Donation>0)
Empir. expectation (refugees)	-0.877**	-0.026***
	(0.439)	(0.008)
Constant	97.180***	2.252***
	(23.729)	(0.304)
Individual controls	Yes	Yes
Observations	1,159	1,159

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.
Results of the second stage regressions of the two-stage least square regression, where the empirical expectation regarding evaluation of the impact of refugees from the Middle East (denoted Empir. expectation (refugees)) is instrumented by the treatment variable (Negative treatment). In column (1) the dependent variable equals the amount that participants dedicated to donate to the UNHCR. In column (2) the dependent variable is an indicator variable that takes value one if the dedicated amount to be donated is larger than zero, and zero otherwise. Individual controls include age, gender, equivalent household income tertile and indication of tertiary education. All regressions include fixed effects of the federal state of residence in Germany and region of participants' (parental) region of origin. Standard errors are clustered on the level of participants' (parental) region of origin. Individual controls include age, gender, equivalent household income tertile and indication of tertiary education.

2.3.4 Indirect reciprocity

Another reason behind the effect that receiving evaluation on own (parental) region of origin had on support for refugees might be the upstream indirect reciprocity. Upstream indirect reciprocity designates a tendency of individuals to exhibit prosocial behaviour towards others because somebody else has exhibited prosocial behaviour towards them (Alexander (1987), Nowak and Sigmund (2005)). Previous studies have provided evidence for the upstream indirect reciprocity, both in the laboratory (Greiner and Levati (2005)) and in the field experiments (Mujcic and Leibbrandt (2018)). Exhibiting

upstream indirect reciprocity in the context of inter-minorities relations would suggest that receiving a less (more) favorable evaluation from an out-group might translate into a less (more) favorable view of another out-group. Thus we would expect more reciprocal participants to react more negatively (positively) in terms of their support for refugees if they were assigned to the Negative treatment (Positive treatment).

In order to test this prediction, a measure of indirect upstream reciprocity was collected using an extended dictator game, whereby one participant (player A) can send a certain sum to another participant (player B), who in turn can send some share of the received amount to a third participant (player C). The amount sent by participant A is multiplied by a factor, which can take either a high or a low value, but the realization of this value is not known to any of the players. Thus, player B observes only the resulting sum they received but is not aware whether it resulted from player A sending a higher sum that was multiplied by a low factor value, or from player A sending a lower sum that was multiplied by a high factor value. Player B is then asked to decide for both scenarios how much of the received sum they would like to send to person C. To ensure that welfare concerns do not play a role in the decision of player B, the amount sent to player C is paid to them without multiplication. Each participant is matched to one of the three roles, and a randomly selected triplet is paid out the amounts according to the decisions they made.

I take the difference in the amount sent to player C in the scenario where player A was more generous versus that when they were less generous as a measure of indirect upstream reciprocity of player B. In order to collect this measure for as many participants as possible, most of the participants were assigned the role of player B ($n = 1150$), and the rest was distributed among the other two roles. All participants assigned to role B received a total of 32 euros (corresponding to player A sending either 8 or 16 euros, and the factor being equal to either 4 or 2, respectively). On average, participants sent 1.21 euros more to player C when player A sent them a higher amount compared to when they sent a lower amount (average amounts sent in two cases was 13.99 and 12.79 euros). This difference is significant (Wilcoxon signed-rank test: $z = 9.544$, $p < 0.001$), providing evidence for behavior consistent with indirect upstream reciprocity.

Furthermore, the distribution of the measure of indirect reciprocity does not differ among treatments (Kolmogorov–Smirnov test: $p = 0.785$), supporting the view of reciprocity as a basic preference.

Table 2.5. The role of upstream indirect reciprocity

	(1)	(2)
	Donation	Pr(Donation>0)
Negative treatment	-5.815***	-0.139*
	(1.557)	(0.073)
Ind. reciprocity	0.433	0.009
	(0.267)	(0.006)
Negative treatment*Ind. reciprocity	-0.751*	-0.033**
	(0.446)	(0.015)
Constant	46.482***	1.002***
	(3.089)	(0.152)
Individual controls	No	No
Observations	1,150	1,150

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Robust standard errors in parentheses.
Column (1) shows Tobit regression of amount dedicated to donate to the UNHCR on the treatment variable, measure of upstream indirect reciprocity (denoted Ind. reciprocity) and their interaction. Column (2) shows the results of Probit regression of the dummy variable that takes value one if participant pledged to donate a value larger than zero on the same set of regressors. All regressions include fixed effects of the federal state of residence in Germany and region of participants' (parental) region of origin. Standard errors are clustered on the level of participants' (parental) region of origin.

Table 2.5 provides the results of the regression of donated amount and that of the dummy variable indicating that participant made a positive donation on the measure of treatment variable, indirect reciprocity, and their interaction. The results indicate that indirect reciprocity indeed had some role in determining the decision to donate. Whereas indirect reciprocity in the positive treatment increased, albeit insignificantly, the pledged donation ($coef. = 0.433$, $p = 0.105$) and the probability to donate ($coef. = 0.009$, $p = 0.167$), it significantly reduced both values in the negative treatment. However, although providing some evidence for the role of indirect upstream reciprocity, these effects are relatively small and do not provide a systematic explanation of the found treatment effects (the treatment variable remains significant).

2.3.5 Preference Falsification

Previous subsections aimed at describing how exposure to expressed prejudice shapes immigration attitudes of individuals with immigration background when these attitudes are expressed privately, that is, when they are unobservable to others (other than the experimenter). However, a broad range of political behaviors, such as protesting, signing a petition, or publically expressing political views, are per construction observable to other members of the polity and, as such, are susceptible to social effects. In particular, due to perceived social pressure, individuals with counter-normative views may prefer to falsify them under observation (Kuran (1997)), such that expressed preferences might not always fully match privately held ones. Previous empirical works convincingly demonstrated that individuals care for how they are perceived by others, and that reputational concerns consequently shape observable behavior in a variety of settings, including political behavior (Valentim (2022), Bursztyn et al. (2020), Enikolopov et al. (2020), DellaVigna et al. (2016), Gerber et al. (2008)).

Whereas most previous works provided evidence of preference falsification in settings where individuals face strongly established norms on pro-social behavior, the context observed here is further complicated by the fact that questions of immigration and asylum policies proved to be highly polarizing among the majority population in (among others) Germany. As social polarization blurs the social consensus on desirable behavior, it is not clear in which direction (if at all) established immigrants might skew their expressed preferences when observed by the majority population.

Understanding how perceived social pressure in the host society might impact immigration attitudes of established immigrants is important not only because preference falsification might mask their genuine preferences but also in light of the findings that expressed controversial preferences, such as xenophobia, might have far-reaching spillover effects, and in the extreme even lead to unraveling of norms that protected against them (Bursztyn et al. (2020)).

In this section, I analyze whether established immigrants feel prompted to express attitudes towards refugees differently when these attitudes can be observed by the

majority population. To get some insight into this, one of the questions that were used in collecting attitudinal measure of support for refugees (q6) was asked again later in the survey, but participants were this time informed that their answer might (or might not) be shown to a participant in a future study. Participants knew that, if used, their response would be provided to a future participant in anonymized form, along with the indication of whether the participant has a migration background and, if so, from which country (countries), and that the person observing their answer would be a German citizen. The question asked participants to rate whether refugees who obtain asylum in Germany make Germany a worse or a better place to live. Participants answered by selecting a number on an 11-points number line, where 0 was indicated as "worse place to live", and 10 as "better place to live". Note that participants were given the opportunity to provide a neutral answer by selecting 5 on the number line, which is exactly in the middle between the two extremes.

I denote the two scenarios as "private" and "observable"[5], and the answers provided in both scenarios by a_p and a_o respectively (note that higher answer indicates a more supportive attitude towards refugees). To compare the answers provided in the two scenarios, I construct a variable $\Delta_o = a_o - a_p$, capturing the extra support that participants expressed in the observable scenario relative to that in the private scenario.[6]

The upper panel of Figure 2.4 shows the average value of Δ_o over a_p. The figure indicates that the average difference in answers strongly depends on the value of the initially provided answer in the private scenario. In particular, participants who expressed less support in the private scenario (provided any answer up to the neutral point (5)), on average, provided systematically higher answers in the observable scenario. More interestingly, participants who in the private scenario indicated highly

[5]The use of the terms "observable" and "private" here is intended only to designate and make easier the distinction between the two scenarios. The ability of the researchers to observe participants' answers renders the private setting clearly distinct from a truly private setting.

[6]I argue that calculating a difference, in this case, is appropriate as the question explicitly asked the participants to rate refugees' influence on a visibly enumerated line, with only endpoints carrying the (exactly opposite) labels. As the answer options are number values (rather than statements, as would be the case in standard Likert scale with different levels of agreement), collected answers can be considered as interval data.

supportive attitudes ($a_p > 7$) systematically decreased their answers in the observable scenario. This suggests that established immigrants, when given an opportunity to misrepresent their attitudes in front of the majority population, do not only use it so as to present themselves as more tolerant than they are , but also to present themselves as less tolerant than they truly are.

One concern here is that the presented evidence of mean reversion when comparing answers in private and observable scenarios might have also resulted if the participants randomly selected their answers in both cases. However, distributions of answers in both scenarios significantly differs from the uniform distribution (Kolmogorov-Smirnov tests for $a_p = U(0, 10)$, and for $a_o = U(0, 10)$, both reject the null hypothesis with $p < 0.001$). Furthermore, as evident from the lower panel of Figure 2.4, the observed degree of preference falsification is significantly lower than the one expected if participants had answered randomly in both scenarios. Nevertheless, this does not exclude the possibility that some share of participants randomly selected their answers, and the others tended not to falsify. However, differently than what would be expected in this case, the distance between the observed and theoretically expected falsification is not equally distributed across the whole range of a_p. Instead, the distance is significantly larger (observed falsification is lower than predicted) among those participants who privately indicated supportive attitudes ($a_p > 5$), than among those who indicated critical attitudes ($a_p < 5$). Additionally, Figure 2.5 depicts the share of participants who falsified upwards ($\Delta_o > 0$) in the upper panel, and the share of those who falsified downwards ($\Delta_o < 0$) in the lower panel, over a_p. As evident from the figure, the observed probability of falsification in both directions discontinuously changes around the neutral position indicated privately ($a_p = 5$). This all suggests that the falsification was rather driven by the perceived social appropriateness of expressed views than by a random behavior.

Finally, the results depicted in Table 2.6, illustrate the effect of experimentally induced status on preference falsification. The table shows the results of an OLS regression of the measured preference falsification (Δ_o) on the treatment variable while controlling for the privately expressed preference (a_p) and a set of individual characteristics. In order

to account for the heterogeneous response to treatment across the distribution of the privately expressed preference, I run the regression separately for participants expressing different levels of support in the private scenario. Specifically, columns (1), (2), and (3) include participants who, in the private scenario, chose an answer that indicates (increasingly) more critical view than the one that would be indicated by selecting a neutral point at $a_p = 5$. Accordingly, columns (4), (5), and (6) include participants who privately indicated (increasingly) more supportive attitudes.

Table 2.6. Treatment effects: Preference falsification

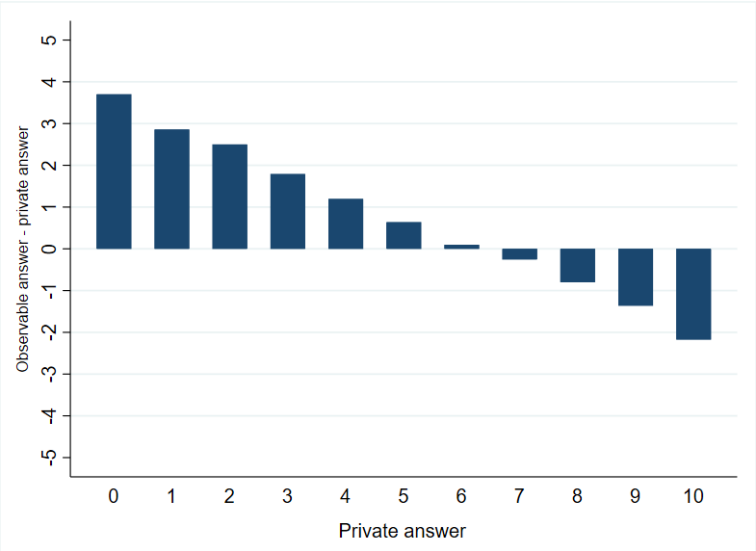
	(1) $a_p < 5$	(2) $a_p < 4$	(3) $a_p < 3$	(4) $a_p > 5$	(5) $a_p > 6$	(6) $a_p > 7$
			$\Delta_o = a_o - a_p$			
Negative treatment	-0.435*	-0.661**	-1.154***	-0.063	-0.165	-0.157
	(0.219)	(0.234)	(0.283)	(0.232)	(0.285)	(0.431)
a_p	-0.650***	-0.667***	-0.720**	-0.594***	-0.690***	-0.826***
	(0.104)	(0.137)	(0.276)	(0.082)	(0.124)	(0.164)
Constant	4.088***	4.050***	5.098***	4.102***	4.985***	5.821***
	(0.267)	(0.403)	(1.097)	(0.913)	(0.987)	(1.487)
Individual controls	Yes	Yes	Yes	Yes	Yes	Yes
Observations	394	288	184	510	395	265

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. OLS regression of measure of preference falsification $\Delta_o = a_o - a_p$, on the privately provided answer a_p and treatment variable. Columns (1) through (3) include only those participants who in private scenario chose an answer that indicates (increasingly) more critical view than the one that would be indicated by selecting a neutral point at $a_p = 5$, as indicated in the columns' title line. Conversely, columns (4) through (6) include only those participants who in private scenario chose an answer that indicates (increasingly) more supportive view than the one that would be indicated by selecting a neutral point at $a_p = 5$. All regressions include fixed effects of the federal state of residence in Germany and region of participants' (parental) region of origin. Standard errors are clustered on the level of participants' (parental) region of origin.

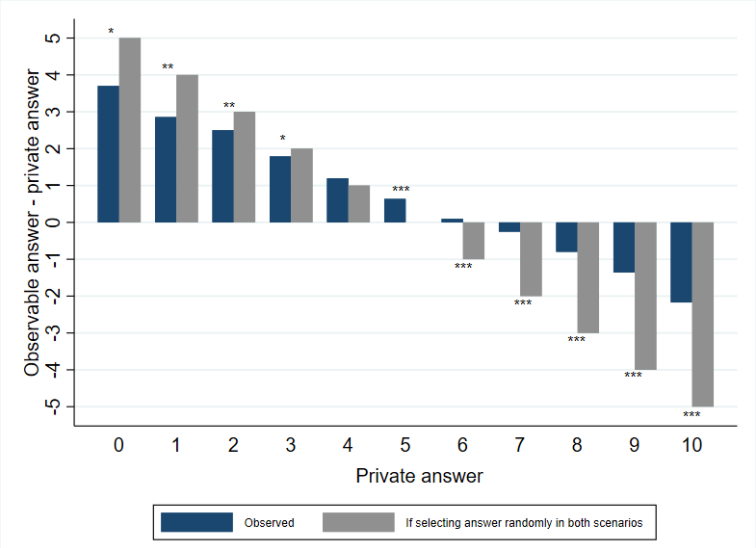
The results show that, among participants who privately indicated more critical attitudes ($a_p < 5$, column (1)), being allocated to the Negative treatment systematically reduced preference falsification. In other words, whereas critical participants falsify their attitudes so as to appear more tolerant in both treatments, those allocated to the Negative treatment do so significantly less. The treatment effect increases in size and precision among those who expressed even more critical views privately ($a_p < 4$, column (2) and $a_p < 3$, column (3)). On the other hand, assignment to the Negative treatment (while still having a negative sign) did not significantly affect preference falsification

among those who privately expressed attitudes that are more supportive than that indicated by the neutral point (i.e., for whom $a_p > 5$), neither when observed together (column (4)), nor when focusing only on those with particularly supportive views (column (5) and column (6)).

These results suggest that expressed prejudice not only negatively affects privately held attitudes towards refugees of those exposed to it, but also increases the readiness to publicly present biased views, thereby weakening the effect of the social norm against xenophobic expressions.

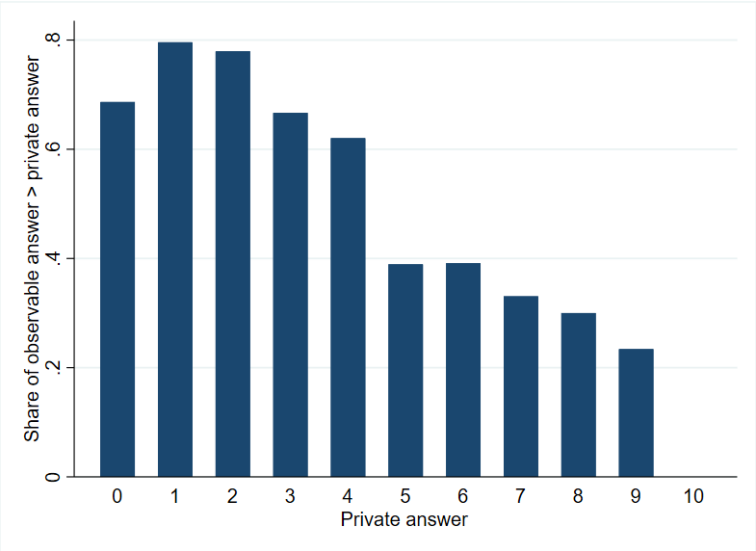


(a)

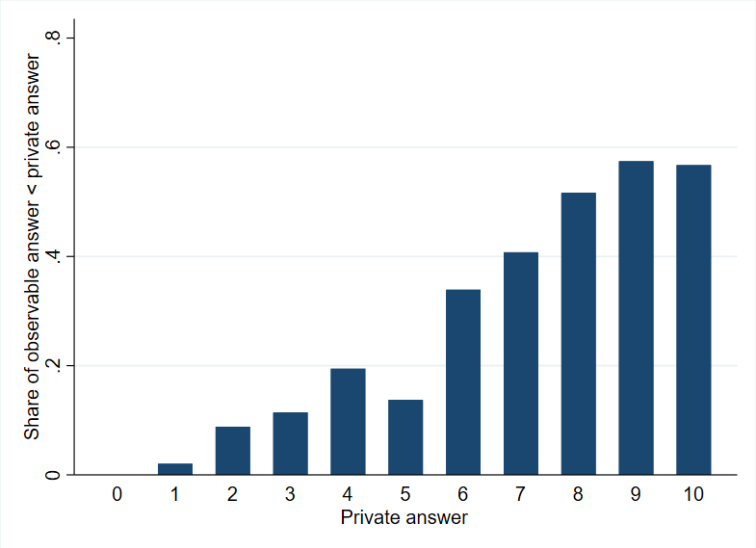


(b)

Figure 2.4. Difference in expressed attitudes - observable v.s. private scenario
The upper panel depicts the average difference between the answer provided in "observable" scenario and the answer provided in "private" scenario. The average difference between the answers is depicted per answer provided in the private scenario. A positive (negative) value indicates that on average participants provided an answer implying more (less) supportive attitude towards refugees when their answer will possibly be observed by a future participant (German citizen), than when answering privately. Note: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$ in sign test for $H_0 : median(\Delta_o) = 0$
The lower panel adds Δ_o that would be expected if both a_p and a_o were selected randomly. Note: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$ in sign test of equal median of observed (Δ_o) and the one that would result under random selection.



(a)



(b)

Figure 2.5. Share of participants with positive and negative preference falsification

The upper panel depicts the average difference between the answer provided in "observable" scenario and the answer provided in "private" scenario. The average difference between the answers is depicted per answer provided in the private scenario. A positive (negative) value indicates that on average participants provided an answer implying more (less) supportive attitude towards refugees when their answer will possibly be observed by a future participant (German citizen), than when answering privately. Note: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$ in sign test for $H_0 : median(\Delta_o) = 0$

The lower panel adds Δ_o that would be expected if both a_p and a_o were selected randomly. Note: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$ in sign test of equal median of observed (Δ_o) and the one that would result under random selection.

2.4 Conclusion

While immigration attitudes received much attention in both economics political science literature, previous research predominantly examined these positions from the point of view of the majority populations of receiving countries. This paper studies the immigration attitudes of established immigrants, that is, those who already have resided in the host countries for a longer time, toward new flows of immigration and the drivers behind these positions. Starting from the implications of the Social Identity Theory, I hypothesize that relative status deprivation, that is, the negative difference in status between own ethnic/national group and that of the native majority (or other, more favorably perceived minorities), has a negative impact on group's members' attitudes toward an even lower ranked status group (e.g., such as refugees). I argue that low-status groups that were socialized in a steep ethnic hierarchy and were exposed to prejudiced treatment, over time, come to perceive ethnic competition as usual and perhaps legitimate, and consequently engage in it also when they are faced with an even lower-status group.

In an online experiment, a sample of participants with immigration background residing in Germany is randomly assigned to receive either a positive or a negative evaluation of the influence of their own (immigrant) in-group on "socio-economic and cultural life in Germany", as expressed by a participant from majority population (without immigration background). Participants are additionally provided with the evaluations of two other out-groups (same for all participants), which fixes the status hierarchy and only leaves the position of the participant's in-group variable. Experimental results confirm the hypothesis by showing that participants who received a negative evaluation of their in-group are significantly less willing to pledge a donation to the UNHCR, and provide less supportive answers to a set of questions regarding attitudes towards refugees (albeit the latter difference is only partially significant).

Furthermore, I hypothesize that the effects of the prejudiced evaluation work through manipulating the social norms surrounding discrimination and its expressions. In particular, people from low-status regions could learn from discrimination directed towards their own in-group that discriminating downwards (i.e., against groups ranked

lower than one's own group) is a widespread behavior in the host society, which in turn increases the probability of them engaging in such behaviors themselves. The results show that, when asked to guess how participants from the native majority evaluated the impact of other immigrant groups, participants who received a negative evaluation of their own in-group (compared to those who received a positive one) expect the evaluations to be significantly more critical. This applies to the expected evaluations of all (mentioned) low-status immigrant groups, including the refugees from the Middle East, but not to the evaluation of a high-status immigrant group. I provide tentative evidence for the role of perceived descriptive norm regarding the acceptance of refugees in mediating treatment effect on behavior.

Lastly, I show that receiving a negative evaluation of the in-group increases the readiness of those participants who privately hold the most negative attitudes towards refugees to publicly state their views, thus weakening the effect of the norm against xenophobic expressions.

The findings presented in this work show how factors specific to the receiving rather than sending country might impact immigrants' political views and behavior. They highlight the importance of policies and public attitudes affecting perceptions of immigrant groups' status, and particularly those seeking to regulate prejudice expressions, by showing how status effects spill over into attitudes towards other, potentially not yet present minorities.

2.5 Appendix

2.5.1 Attitudinal questions on views regarding refugees from the Middle East

Participants were asked to provide answers to the following seven questions. Other than the question number 6, all questions have been adopted from Dinas et al. (2021a).

1. Do you think Germany should increase or decrease the number of people it grants asylum to? (1 = Greatly increase; 5 = Greatly decrease)
2. Refugees are a burden on our country because they take our jobs and social benefits.(1 = Completely agree; 5 = Completely disagree)
3. The money spent on the accommodation of refugees in our country could have been spent better to cover the needs of Germans. (1 = Completely agree; 5 = Completely disagree)
4. Refugees will increase the likelihood of a terrorist attack in our country. (1 =Completely agree; 5 = Completely disagree)
5. Refugees in our country are more to blame for crime than other groups. (1 =Completely agree; 5 = Completely disagree)
6. Is Germany made a worse or a better place to live by refugees who are granted asylum in Germany? (Respondents select their answer on a enumerated scale, where value 0 is labeled as "Worse place to live", and value 10 is labeled as "Better place to live")
7. Among the following options, which one do you think best explains why refugees from Syria and other countries leave their country? (1 = To flee war; 2 = To improve their economic conditions; 3 = To avoid political persecution; 4 = To gain access to host country's social benefits.)

2.5.2 Sample description

Tables 2.8 shows the basic demographic characteristics of the sample as a whole, and separately for both treatments. Table 2.7 shows the distribution of the sample across the targeted regions of origin.

Table 2.7. Regions of origin per treatment

	Share across treatments		
	Positive treatment	Negative treatment	Total
Region of (parental) origin			
Bulgaria & Romania	0.065	0.078	0.072
Central-Eastern European Union (Czech Republik, Slovakia, Poland, Hungary)	0.182	0.172	0.177
Baltic states (Estonia, Lithuania, Latvia)	0.011	0.007	0.009
Ex-Yugoslavia (Bosnia and Herzegovina, Croatia, Kosovo, Montenegro, North Macedonia, Serbia, Slovenia)	0.058	0.084	0.072
North Africa (Morocco, Algeria, Lybia, Tunesia and Egypt)	0.078	0.068	0.073
Southern European Union countries (Greece, Italy, Portugal, Spain, Cyprus and Malta)	0.106	0.145	0.127
Turkey	0.249	0.205	0.226
Southern Ex-Soviet union (Tajikistan, Turkmenistan, Georgia, Kazakhstan, Kyrgyzstan, Armenia and Azerbaijan)	0.063	0.073	0.068
Western Ex-Soviet union (Ukraine, Moldova, Belarus)	0.033	0.028	0.030
Russian federation	0.092	0.084	0.088
Albania	0.063	0.056	0.060
Observations	554	605	1,159

Notes: Regions of participants' own or parental origin across treatments.

Table 2.8. Sample description

	Means across treatments		
	Positive treatment	Negative treatment	Total
Age			
[18-24]	0.338	0.349	0.343
[25-34]	0.300	0.284	0.292
[35-44]	0.182	0.175	0.179
[45-54]	0.108	0.116	0.112
[55-64]	0.060	0.063	0.061
[65-74]	0.011	0.010	0.010
[75-84]	0.002	0.003	0.003
Gender			
Male	0.457	0.438	0.447
Education			
Primary or lower secondary	0.354	0.331	0.342
Secondary	0.233	0.238	0.236
Tertiary	0.413	0.431	0.423
Equivalised household income			
Tertile 1	0.372	0.367	0.369
Tertile 2	0.361	0.385	0.374
Tertile 3	0.267	0.248	0.257
Observations	554	605	1,159

Notes: Demographic characteristics of the sample per treatment.

2.5.3 The role of participants' mood

The following table replicates the results described in Table 2.1, while controlling for the three measured affective dimensions - pleasure, arousal, and dominance, elicited via Self-Assessment Manikin questionnaire.

Table 2.9. Treatment effects: The role of participants' mood

	(1)	(2)
	Donation	Pr(Donation>0)
Negative treatment	-6.503***	-0.170***
	(1.583)	(0.062)
Valence	1.455***	0.058***
	(0.513)	(0.011)
Arousal	0.906	0.042***
	(0.629)	(0.015)
Dominance	-0.259	-0.023
	(0.818)	(0.022)
Constant	32.758***	0.790**
	(6.289)	(0.318)
Individual controls	Yes	Yes
Observations	1,159	1,159

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.
Column (1) shows Tobit regression of amount dedicated to donate to the UNHCR on treatment variable, elicited measures of participant's mood, and the set of individual controls. Negative treatment indicates receiving negative status information regarding own in-group (with Positive treatment serving as a baseline). Columns (3) shows Probit regression of an indicator variable for donation being larger than zero on treatment variable and the set of individual controls. All regressions include fixed effects of the federal state of residence in Germany. Individual controls include age, gender, equivalent household income tertile and indication of tertiary education. Standard errors are clustered on the level of region of origin.

2.5.4 Survey

The following pages show the transcript of the survey that would be shown to a hypothetical participant who selected Montenegro as their own or parental country of birth.

Instructions

We are a group of scientists from one of the Max-Planck institutes. This survey should last (on average) around 15-20 minutes. Your answers will be available only to the researchers, in anonymized form and there will be no possibility to identify the respondents. The data will be kept on our servers and will be treated confidentially. Anonymized data can be provided to other researchers. You can decide at any point of time to leave the survey.

Please select whether you would like to participate in this survey.

- ○ Yes, I would like to participate in this survey
- ○ No, I would not like to participate in this survey

Q: Are you male of female?

- ○ Female
- ○ Male

Page Break

Q:
Do you speak or understand a language other than German?
Please select all that apply.

Names of languages in the list are in English.

- ☐ *No, I speak and understand only German*
- ☐ ALBANIAN
- ☐ ARABIC

Note: a total of 79 languages were listed

Q:
What is the country of your birth?
Please select from the list below.

Names of countries in the list are in English

▼ Afghanistan ... Zimbabwe

Page Break

Q:
What is the country of birth of your **mother**?
Please select from the list below.

Names of countries in the list are in English

▼ Afghanistan ... Zimbabwe

Page Break

Q:
What is the country of birth of your **father**?
Please select from the list below.

Names of countries in the list are in English

▼ Afghanistan ... Zimbabwe

Page Break

Q: How old are you?

Q:
In one previous study, 125 German participants (without migration background) were asked to evaluate the impact of immigrants coming to live in Germany from different regions/countries on **socio-economic and cultural life in Germany**.

Participants were asked to evaluate the impact of several immigrant groups.

On the next screen, you will be shown a subset of collected answers.
Please read these answers carefully.

Note: If participant was matched to region of Ex-Yugoslavia and was allocated to the Positive treatment
Q:
If participant was matched to region of Ex-Yugoslavia and was allocated to the Negative treatment:

People that come to Germany from countries of the **Ex-Yugoslavia**:

Croatia, Bosnia and Herzegovina, Slovenia, Serbia, Kosovo, Montenegro und North Macedonia

Contribute rather **NEGATIVELY** to the socio-economic and cultural life in Germany

People that come to Germany from countries of the **western EU**:

Austria, Belgium, France, Ireland, Luxemburg, Netherlands

Contribute rather **POSITIVELY** to the socio-economic and cultural life in Germany

People that come to Germany from **Lebanon**:

Contribute rather **NEGATIVELY** to the socio-economic and cultural life in Germany

Note: If participant was matched to region of Ex-Yugoslavia and was allocated to the Positive treatment

Q:

People that come to Germany from countries of the **Ex-Yugoslavia**:

Croatia, Bosnia and Herzegovina, Slovenia, Serbia, Kosovo, Montenegro und North Macedonia

Contribute rather **POSITIVELY** to the socio-economic and cultural life in Germany

People that come to Germany from countries of the **western EU**:

Austria, Belgium, France, Ireland, Luxemburg, Netherlands

Contribute rather **POSITIVELY** to the socio-economic and cultural life in Germany

People that come to Germany from **Lebanon**:

Contribute rather **NEGATIVELY** to the socio-economic and cultural life in Germany

Q: To which degree do you agree with the stated results?

	Strongly agree	Agree	Neither agree nor disagree	Disagree	Strongly disagree
	1	2	3	4	5
					

Q: We would now like to ask **your opinion regarding immigration to Germany**. Particularly, we would like to ask your opinion about today's asylum seekers in Germany.

Page Break

Q: Do you think Germany should increase or decrease the number of people it grants asylum to?

	Greatly increase	**Increase**	**Neither increase nor decrease**	**Decrease**	**Greatly decrease**
	1	2	3	4	5
Indicate your opinion					

Page Break

Q: Refugees are a burden on our country because they take our jobs and social benefits.

	Strongly agree	**Agree**	**Neither agree nor disagree**	**Disagree**	**Strongly disagree**
	1	2	3	4	5
					

Page Break

Q: The money spent on the accommodation of refugees in our country could have been spent better to cover the needs of Germans.

Strongly agree	Agree	Neither agree nor disagree	Disagree	Strongly disagree
1	2	3	4	5

Page Break

Q: Refugees will increase the likelihood of a terrorist attack in our country.

Strongly agree	Agree	Neither agree nor disagree	Disagree	Strongly disagree
1	2	3	4	5

Page Break

Q: Refugees in our country are more to blame for crime than other groups.

Strongly agree	Agree	Neither agree nor disagree	Disagree	Strongly disagree
1	2	3	4	5

Q: Among the following options, which one do you think best explains why refugees from Syria and other countries leave their country?

- ○ To flee war
- ○ To improve their economic conditions
- ○ To avoid political persecution
- ○ To gain access to host country's social benefits.

Page Break

Q: Is Germany made a worse or a better place to live by refugees who are granted asylum in Germany?

Worse place to live **Better place to live**

0 1 2 3 4 5 6 7 8 9 10

Q: As a part of this study, we will run a lottery. All participants who complete the survey will automatically be included as participants in the lottery and everyone has the same chance to win. The winner of the lottery will receive a prize in value of **100 Euros**, which will be paid out in addition to their other earnings from this survey to their panel account.

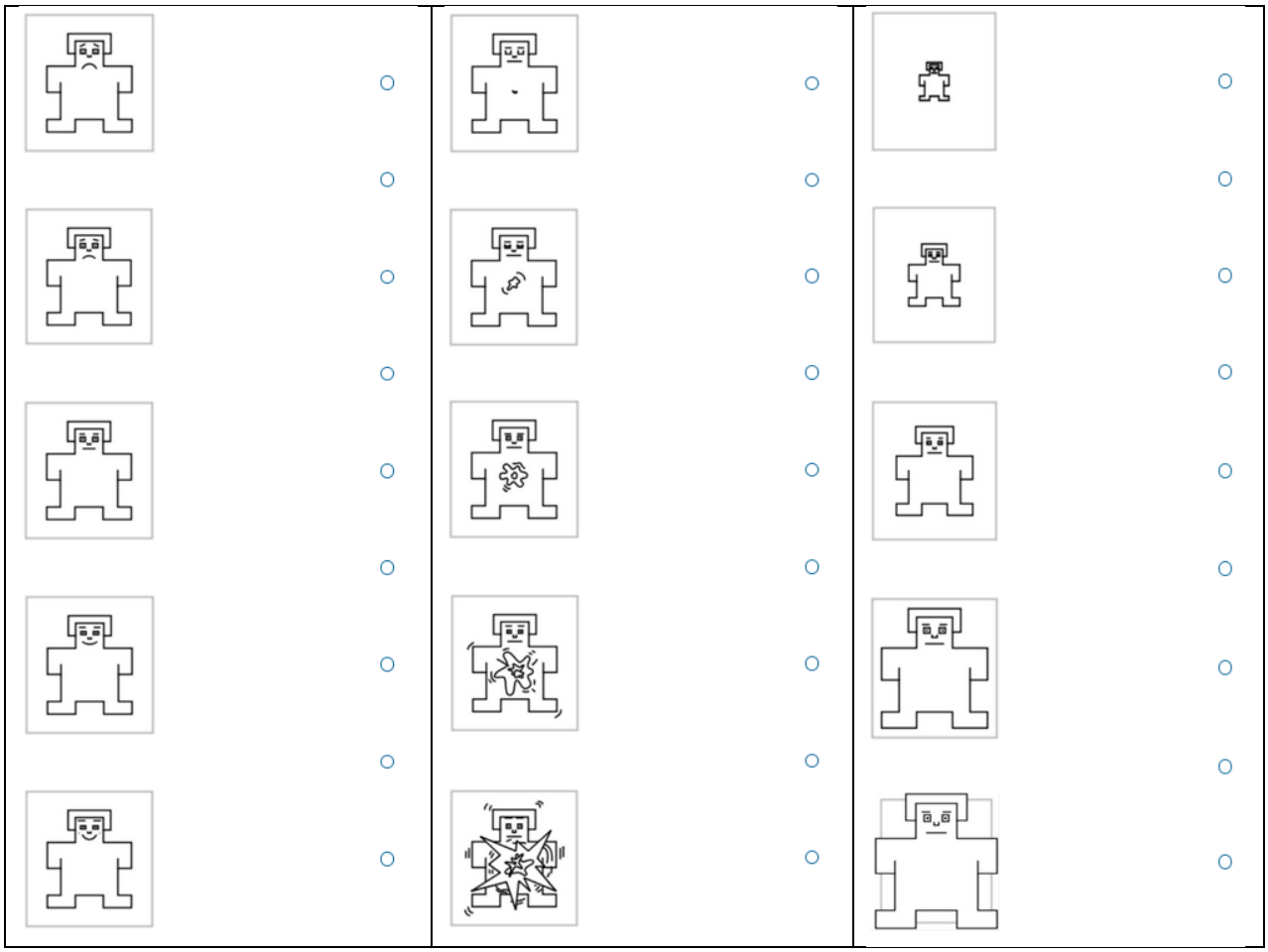
However, you can also choose to donate one part of the lottery prize that you might win to the United Nations High Commissioner of Refugees (UNHCR), a global organization dedicated to helping refugees. If you win the lottery, the donation amount will be automatically deducted from the 100-EUR prize, and a donation in this value will be made to UNHCR. The remaining part of the 100-EUR prize will be paid out to you.

Please indicate below if you would like to donate some part of your 100-EUR prize in case you win the lottery, and if so, how much.

If I win the lottery, I would like to donate the following amount from my 100-EUR prize
(Please enter an amount in EUR, between 0 and 100)

_____

Q: Please select in each of the three sections (boxes) below the figure that describes your current mood the best. If you feel that your mood lies in between two figures, please select the point between them.

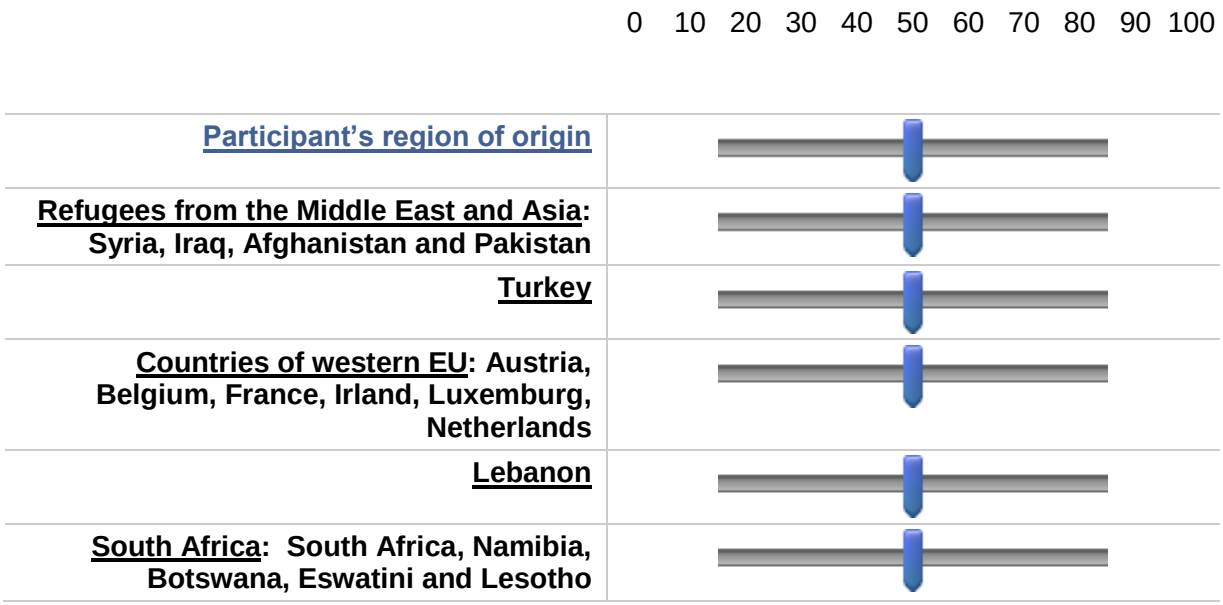


Q: At the beginning of the survey, we have told you about the earlier study, where 150 German participants (without migration background) were asked to evaluate the impact of immigrants coming to live in Germany from different regions/countries on socio-economic and cultural life in Germany. In that study participants were asked for each of several countries/regions to evaluate whether people coming to Germany from this country/region:

A: "contribute rather POSITIVELY to socio-economic and cultural life in Germany", or
B: "contribute rather NEGATIVELY to socio-economic and cultural life in Germany"
You have been shown some of the collected answers.

Can you try to guess **how many percent of all asked participants evaluated NEGATIVELY socio-economic and cultural impact of people immigrating to Germany from each of the following countries/regions?** (In other words, what percent of participants selected option B)

The respondent whose answers are the closest to the true outcome collected in that study which will be shown on the next screen) will receive an additional **25 EUR** to their panel account when the survey is completed.



Indirect reciprocity if participant is matched to the role of „Person B"

Q: In this part of the study, you will interact with two other participants of this study, but you will not know who these participants are. Let us call the three participants person A, person B and person C. **You are person B**.

The rules are the following:
Person A is given a budget of **30 EUR**. Person A can decide to send from this budget to person B (i.e. to you) some amount. The amount that person A sends to you will then be multiplied by a **factor**, and the resulting amount will be paid out to you. The value of the factor is not known to any of the three persons, but it can either be equal to 2 or 4.
Person A will keep for themselves the rest of the budget, that is the amount that they did not send to you.

------- *Example* ---------
For example, if person A sends you 5 EUR, then persopn A keeps for them 25 EUR and you will receive:
*If **factor=2**: you will receive 5 EUR***2**=10 EUR*
*If **factor=4**: you will receive 5 EUR***4**=20 EUR*

You will know the total amount that you received from person A. However, you will not know the value of the factor by which the amount that person A sent to you was multiplied.
After you learn the total amount that you receive from person A, you can decide whether you want to send some part of that amount to person C, and if so, how much. This amount will not be multiplied by the factor. The amount that you send is paid out to person C. You will keep the rest (that is, the part that you do not send to person C) for yourself.

All participants in this study are assigned one of the roles (person A, person B or person C) and matched into groups of three participants. When all participants complete the survey, one group will be randomly selected and, taking into account their decisions (how much person A sent to person B, and person B to person C), the payments will be made to these participants. Person A with whom you are matched has already completed the survey and decided how much to send to you. Please proceed to the next screen to learn about the amount that you receive and to make the decision regarding the amount that you want to send to person C.

Q:

You received the amount of 32 EUR

As stated on the previous page, you don't know what was the value of the multiplication factor. That means that one of the following scenarios happened:

Scenario1: Person A sent you **8 EUR** and the factor equals **4**. This is why you received **8 EUR***4=32 EUR.

Scenario2: Person A sent you **16 EUR** and the factor equals **2**. This is why you received **16 EUR***2=32 EUR

For each of the two scenarios, please make a decision how much of the received 32 EUR (if any) would you like to send to person C.

Scenario 1:

If person A sent me **8 EUR**, and the **factor=4**, I would like to send to person C the following amount:

	0	32
Please select in EUR		

Scenario 2:

If person A sent me **16 EUR**, and the **factor=2**, I would like to send to person C the following amount:

	0	32
Please select in EUR		

Indirect reciprocity if participant is matched to the role of „Person A"

Q: In this part of the study, you will interact with two other participants of this study, but you will not know who these participants are. Let us call the three participants person A, person B and person C. **You are person A**.

The rules are the following:
Person A (that is, you) is given a budget of **30 EUR**.
You can decide to send from this budget some amount to person B (person B will not receive the budget). The amount that you send to person B will then be multiplied by a **factor**, and the resulting amount will be paid out to person B. The value of the factor is not known to any of the three persons, but it can either be equal to 2 or 4.
You will keep for yourself the rest of the budget, that is the amount that you don't send to person B.

------- *Example* ---------
For example, if you send 5 EUR to person B, then you keep for yourself 25 EUR and person B will receive:
*If **factor=2**: you will receive **5EUR*2=10 EUR***
*If **factor=4**: you will receive **5EUR*4=20 EUR***

Person B will know the total amount that they received from you. However, they will not know the value of the factor by which the amount that you sent was multiplied. Person B will then be asked if they want to send some amount from the received sum to person C. The amount that person B sends to person C will not be multiplied by the factor. Instead, it will simply be paid out to person C.

All participants in this study are assigned one of the roles (person A, person B or person C) and matched into groups of three participants. When all participants complete the survey, one group will be randomly selected and, taking into account their decisions (how much person A sent to person B, and person B to person C), the payments will be made to these participants.

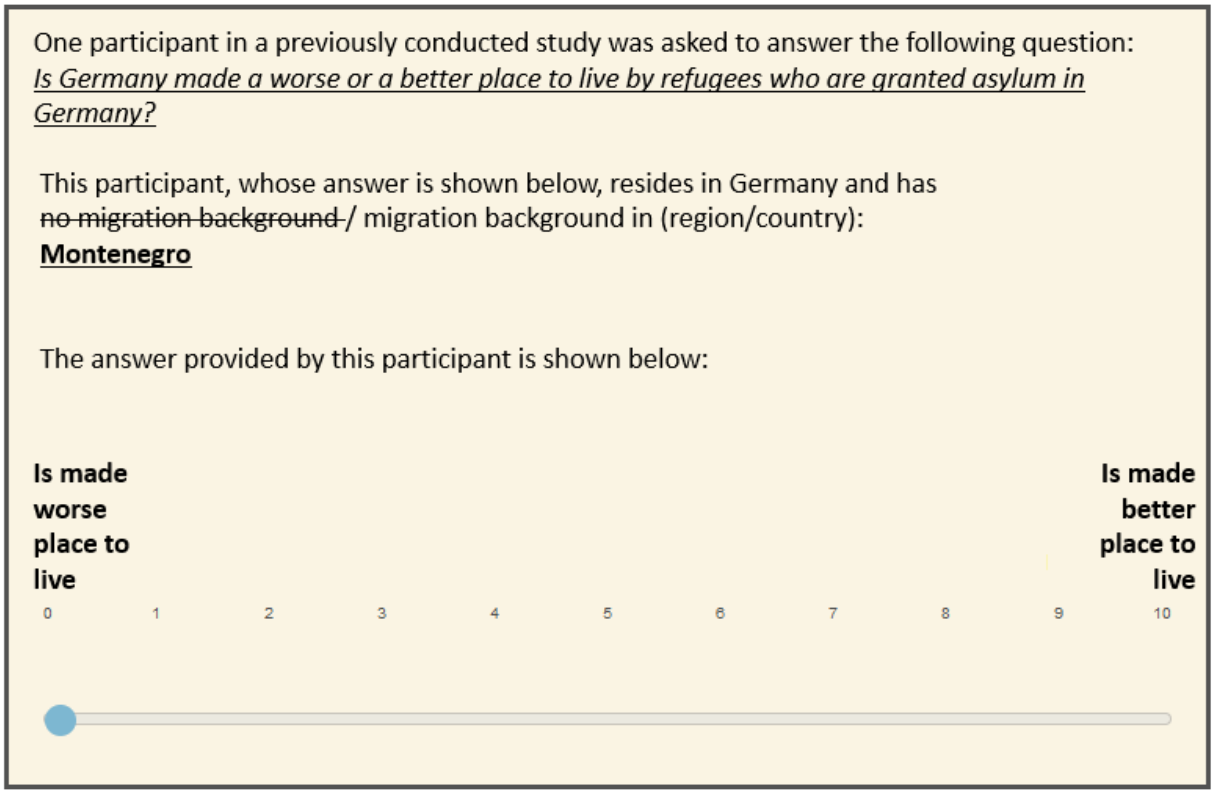
Please indicate below (in EUR), how much would you like to send to person B from your 30 EUR budget.

- ○ 0 EUR
- ○ 8 EUR
- ○ 16 EUR
- ○ 30 EUR

Q:
Please note that all the answers that you provided so far will be seen (in anonymized form) only by the researchers. **On this page, we would like to ask you to provide an answer to the question: *"Is Germany made a worse or a better place to live by refugees who are granted asylum in Germany?"*, that we can show to participants in a study that might be conducted in future**.

If conducted, this study will be run in Germany, and the participants will be citizens of Germany. Your answer in anonymized form would be shown to participants to inform them on your views. Particularly, participants will be shown only the form shown below with the answer that you choose.
Please complete the form with your answer to the stated question.

Note: An example of the form that would be shown to a participant matched to region of origin ex-Yugoslavia, and particularly who indicated Montenegro as own (parental) country of birth.



Q: On the next page, you will be asked to answer one question. The question is aimed to measure numerical ability.
You will have 1 minute to provide the answer to this question. There are 6 offered answers, and only 1 of those is correct.
The participant who provides the **correct answer in the shortest time** will be paid **30 EUR** in addition to other earnings from this study.

Page Break

Q: A bat and ball cost $1.10. The bat costs one dollar more than the ball. How much does the ball cost?
Please enter your answer below in cents.

Page Break

Q: Before you proceed, we would like to ask your opinion on the question that you have seen. We might use the same question in a future study to measure the numerical ability of other participants. Any participant who took part in this survey would be excluded from the future survey.
If we decide to use this question, participants in that study will be provided as here 60 seconds, to answer the questions.

We would like to ask your opinion: If we decide to use this question in the future study, do you think that we should reward by 30 EUR not only the fastest correct answer, but also the second- and/or third-fastest correct answer? Alternatively, do you think that 30 EUR is a too high prize for this task and should be reduced to 20 EUR?

We will collect the opinions of all participants in this study. The option that receives the most votes will be implemented.

Please vote below.

- ○ Reward **only the fastest** correct answer by **30 EUR**.
- ○ Reward **the fastest and second-fastest** correct answer by **30 EUR each**.
- ○ Reward the **fastest, second-fastest and third fastest** correct answer by **30 EUR each**.
- ○ Reward **only the fastest** correct answer by **20 EUR**.

Q: If you were not born in Germany, since when do you live in Germany?

▼ I was born in Germany ... 1950

Page Break

Q: Do you have a German citizenship?

- ○ Yes
- ○ No
- ○ I have a German and other citizenship(s)

Page Break

Q: How many persons live in your household (including yourself):

- ○ Adults _____
- ○ Children (below 18 years old) _____

Q7 What ist he highest level of schooling that you completed (allgemeinbildend oder beruflich)?
Note: the options shown below designate educational levels common in Germany

- ○ Ich habe die Schule vor Ende der 9. Klasse / ohne Hauptschulabschluss verlassen
- ○ Hauptschulabschluss
- ○ Realschulabschluss (Mittlere Reife)
- ○ Hochschulreife / Abschluss des Abiturs (Gymnasium bzw. EOS)
- ○ Lehre / Berufsausbildung
- ○ Bachelor
- ○ Master / Diplom / Magister / Staatsexamen oder vergleichbarer Abschluss
- ○ Promotion

Q: Please select below federal state of your residence and your municipality / city / region:
Federal state
Municipality / city / region

▼ Baden-Württemberg ... Thüringen ~ Weimarer Land

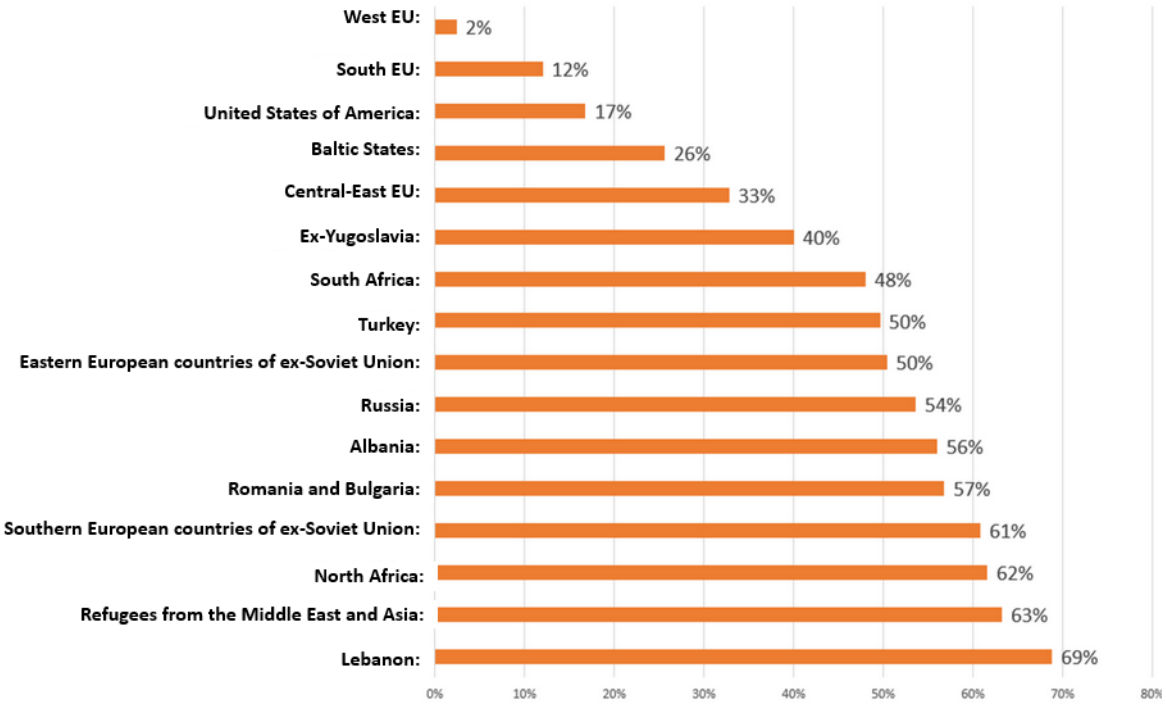
Q: What was the monthly net income of your household in the last month? By monthly net income we mean the sum that you received through wages, labor income, Income from self-employment, retirement income, pension, rent, Income from public subsidies, income from renting, leasing, housing benefit, child benefit and other income **after deducting taxes and social security contributions**.

- ○ < 1000€
- ○ 1000€-1999€
- ○ 2000€-2999€
- ○ 3000€-3999€
- ○ 4000€-4999€
- ○ 5000€-7500€
- ○ > 7500€

Q: As was explained before, you will now be shown the full results of the previously conducted study with 125 participants without immigration background. For each depicted immigrant-group you will be shown the percent of participants who evaluated negatively the impact of this group on socio-economic and cultural life in Germany.

Page Break

Q:
Percent of 125 participants without immigration background who evaluated negatively the impact on socio-economic and cultural life in Germany of immigrants coming to Germany from following regions/countries.



Chapter 3

The influence premium of monetary status

This chapter is based on joint work with Andrea F. M. Martinangeli.

3.1 Introduction

The transmission of adaptively valuable behaviours, central in the cultural evolution of human societies (Fried (1967), Henrich and Gil-White (2001)), relies, on one hand, on status hierarchies assigning prestige and influence to individuals who more likely possess valuable traits and knowledge. On the other, it relies on others' ability to recognise even minimal status and rank cues (Smith (1982), Maner et al. (2008), Shariff et al. (2012), Koski et al. (2015), Witkower et al. (2020)), such that transmitted behaviours are more likely to originate from high-status, more successful individuals (Henrich and Gil-White (2001), Cheng and Tracy (2013)). The question arises, with the accumulation of material status sources (Hill (1984), Cheng and Tracy (2013)), how sensitive the influence enjoyed by high status (material or monetary) individuals over others' behaviours (Walker (2015)) is to the existence of an underlying knowledge or skill-based source for their status.[1] Answering this question entails isolating any impact of status

[1]We restrict our analysis of status exclusively to its economic dimension by varying endowment size as in much of previous experimental research (e.g., Fehr (2018), Martinangeli and Martinsson (2020), Markussen et al. (2021)). Alternative dimensions concern individuals' relation with production processes, their competence or morality, or their access to power (e.g., Koski et al. (2015), Mattan et al. (2017)). Their discussion is beyond the scope of this paper. We will henceforth often use the words "material" or "monetary" in reference to the same conception of status based purely on the possession of greater resources than others.

from both the socio-cultural shared narratives surrounding status itself and from its possessor's individual characteristics.

We do not dispute that individual skill and effort play an enormous role in determining individual success: Our focus is on that component of individuals' standing which cannot be attributed to individual effort or skill, but is instead due to circumstances beyond the individuals' control. For much of European history up to the modern era, for instance, (quasi) randomly assigned birth-right status accorded the aristocracy, often irrespective of skill, enormous privileges and disproportionate influence over others.[2] Even nowadays, intergenerational wealth transmission is shown to determine a large proportion of individuals' economic conditions (e.g., Oulton (1976), Piketty (2011), Elinder et al. (2018), Ohlsson et al. (2020)), and apparent status alone (clothing, vehicle, voice and demeanor) is shown to generate expectations of competence, trustworthiness and cooperativeness, granting its possessor preferential access to resources and influence (e.g., Christopher and Schlenker (2000), Fiske et al. (2002), Nelissen and Meijers (2011), Little and Roberts (2012), Bull and Rumsey (2012), Duarte et al. (2012)).

We thus investigate whether status holds a "premium" in its own right in the beholders' eye, independent of its original evolutionary purpose. To the best of our knowledge, this paper presents the first evidence of the status-influence nexus free of both the cultural (e.g. shared narratives and beliefs about the origin of wealth) and individual (e.g. skill or education) confounds surrounding it (Cook (1975), Ridgeway (1991), Stewart and Moore Jr (1992), Weber (2001), Fehr (2018), Wooldridge (2021)).

We rely on a narrow monetary definition of status, whereby high status is purely determined by holding a large rather than a small endowment, and on an operationalisation of influence allowing for sharp predictions and clean tests in a parsimonious experimental paradigm (Falk and Heckman (2009)). Monetary status has been shown in previous experimental research to profoundly impact own and others'

[2]See Wooldridge (2021) for a discussion.

expectations and behaviours (e.g., Nelissen and Meijers (2011), Kuziemko et al. (2014), Martinangeli and Windsteiger (2020), Martinangeli (2021), Rockenbach et al. (2021)).[3]

Specifically, we observe how decision makers from a representative sample of the German population (with respect to gender, age, income and geographic area), facing a purely individual choice without any externality, are influenced in their decisions by uninterested third-party advice. By orthogonally varying the advisors' status and its informativeness about their underlying cognitive ability, we are able to investigate the relative merit of the two as sources of influence, and to tap into their potentially far reaching consequences on everyday interactions for individuals and organizations. Receiving advice is ubiquitous in decision making. From purchases of goods and services to selection of strategies in social dilemmas (Schotter (2003), Schotter and Sopher (2006), Chaudhuri et al. (2006), Schotter and Sopher (2007)), the effectiveness of advice from even minimally experienced peers in steering individuals' decisions is well documented in research (e.g., Chevalier and Mayzlin (2006), Keller and Fay (2012)). Simply put, a disproportionate influence of advice originating from the advisors' status might well be detrimental should it be orthogonal to their ability to deliver quality advice.

We focus on the vast number of situations in which people rely on the "naïve" word-of-mouth advice, often unsolicited, from friends or workplace peers who are hardly more experienced in the given task than the decision makers themselves (Schotter (2003)). Should advisors (e.g. professional consultants, lawyers, brookers, or senior co-workers) be substantially more experienced in the matter at hand than than the decision makers (advisees, henceforth), and should their expertise be purposefully purchased by the latter,

[3]Because previous studies were not designed with the purpose of identifying the influence premium of status, these observations are confounded by positional preferences, other regarding concerns, beliefs, and normative expectations. Martinangeli (2021) and Rockenbach et al. (2021) find that greater prosociality is expected on behalf of, respectively, subjects with a high endowment and inhabitants of wealthy neighbourhoods. Both studies observe stronger conditional behaviours with respect to the expected or observed behaviours of the rich than of the poor. Other regarding concerns and the material impact that cooperation on behalf of wealthy individuals limit however the informativeness of these studies on their power to induce behaviour-change in others. Nelissen and Meijers (2011) randomises whether charitable donations are solicited by a poorly or lusciously dressed individual. Because the subjects of study (the donors) are unaware of the experimental design, their worldviews, beliefs and culturally acquired narratives about the origin of the individual's status confound the impact of the mere monetary status conveyed by the clothing.

their experience will plausibly (and reassuringly) dominate over other factors potentially determining the influence of their advice, thus impeding the investigation of the role of status (e.g., Gino (2008)).

Advice is moreover particularly important when the decision-maker faces a problem without a clearly correct answer (Brockner et al. (1984)). In these situations, the power of third parties without vested interests play an important role in steering the individuals' choice, and hence in determining the final outcome and the decision maker's ultimate welfare. Where a clear optimal choice exists, the power of advice in driving choices away or towards the optimum will certainly be weakened by the attractiveness of the optimal action, commonly visible to and understood by both the advisor and the decision maker.

On the other hand, where the optimal choice is mostly determined by the decision maker's preferences, the power of advice of different origin in steering choices can be more readily measured because, provided that the advisor has no insight in decision maker's preferences, advice and optima will on average be orthogonal. How much to invest in risky assets (financial, or immaterial goods with uncertain returns like education or interpersonal relationships), how much to contribute to collective efforts or how much trust to place in others or institutions, are all situations in which no unequivocally and universally correct action exists. Despite objective measures of riskiness do exist, the ultimate benchmark against which to evaluate the appropriateness of one choice over another are the decision maker's own and purely individual preferences over risk.

In their simplest and most stylised form the examples mentioned above can be narrowed down to the individual choice of purchasing a lottery. This conceptualisation allows us to shine an even brighter light on the sources of advisors' influence by removing the social dimension from the situation. In contrast, social dilemmas (e.g., Martinangeli (2021)), charitable donations Nelissen and Meijers (2011) or trusting decisions (e.g., Xiao and Bicchieri (2010), Bicchieri et al. (2011)), all allow for an important and potentially confounding role of other regarding concerns, beliefs and social norms in steering individuals' behaviours and outcomes. Moreover, by minimising the power of commonly understood optimal choices, we are able to capture how well advice of different origin is capable of influencing decision maker's choices. We therefore

confront the agent with a perfectly individual choice allowing us to observe how choices are influenced by others free of any welfare externality, commonly understood norms of behaviour or optima which would confound our results.[4]

Concretely, to elicit the pure impact of status on individuals' influence over others, we face individuals with the choice of investing any fraction of an endowment (randomly assigned to high or low) in a lottery game yielding either zero or triple the investment with equal probabilities (Gneezy and Potters (1997)). Before choosing their investment, the participants receive advice from another participant of explicitly high or low endowment status. We can thus clearly attribute any difference in the impact of the advice to the advisors' status: The advisor is commonly known to having faced the same choice under the same conditions, and their endowment (status) was randomly assigned to high or low. Notice moreover that no "correct" investment can be suggested without knowledge of the investor's risk preferences, voiding the advice of any meaning and strengthening the result for any observed differential propensity to follow advice originating from specific advisor types.[5] Finally, notice that our design allows us to separate status from the confounding system of beliefs that accompany it.[6]

Following the evolutionary arguments presented early on in this paragraph, our first hypothesis is therefore that even when status is entirely uninformative about the advisor's underlying characteristics, individuals respond to the status signal by assigning high status advisors an influence premium over low status advisors (Hypothesis 1), i.e. advice from high status advisors is followed to a larger extent.

Further, to place our results in relation to other prominent aspects of this problem, namely the advisor's cognitive ability, we complement this design with one in which,

[4]In this respect, our approach differs from the conceptualisation of influence proposed by Robert Cialdini and associates (e.g., Cialdini (2001)). Influence is there envisaged as individuals' ability to steer others' decisions to obtain a desired result by leveraging on social mechanisms and situational cues: reciprocity, commitment and consistency, social proof, authority, scarcity, liking and social identity, none of which play a role in our framework. Our approach, rather, investigates the influence over others' actions accorded to entirely uninterested others by the mere resources in their possessions.

[5]With this feature our design departs from the setup in Ronayne and Sgroi (2018), who instead study the impact of explicitly good or bad advice

[6]It would not be the case, for instance, were we randomly distributing advice on behalf of poorly or lusciously dressed individuals (for instance, Nelissen and Meijers (2011)). In this case, the dress would carry both its status and its confounding ability significance into the experiment, such that the two could not be disentangled unless the random origin of the dress style is revealed to the study subjects.

instead of being randomly assigned, status is earned by the advisor in a cognitive ability test. This branch of the experiment allows us to capture how advisors' status shapes their power to influence others when granted based on an individual characteristics presumably allowing them to offer "better" (or at least more informed) advice.[7] Because of the added complexity introduced by the cognitive ability component in the status-influence link, we do not and will not make a comparison of the results obtained across the two branches (random and earned endowment) of the experiment. Cognitive ability is a radically different trait from pure monetary status, inducing deep emotional, psychological and attitudinal responses (e.g. status rejection, advice rejection, spite, self-confidence to name only a few) the investigation of which is left out of the scope of this paper.

Our second hypothesis is that when status is informative about underlying cognitive ability, we again expect the influence of high earned status advisors on the decision maker to be greater than that of low earned status advisors (Hypothesis 2).[8]

Our result squarely support our first hypothesis: High status is rewarded with greater influence, in that respondents who received advice from a *high* random-status advisor are significantly closer to the advised investment level than those who received advice from a *low* random-status advisor. We show that this result is driven by individuals (advice receivers) who scored below average in a cognitive ability test. On the other hand, we do not find support for our second hypothesis: High and low earned-status advisors do not impact our respondents' choices differently. This result should be read cautiously: Earned status has a very different meaning compared a randomly assigned one, and this experiment was not designed to control for all mechanisms at play in such scenario.

This article is organised as follows: Section 3.2 describes the experimental design, Section 3.3 illustrates our empirical strategy and our results, Section 3.4 concludes.

[7] Though "good" advice remains here, for all purposes, an entirely subjective matter.

[8] These hypotheses are pre-registered with the AEA RCT repository (registered trial AEARCTR-0007269). We deviate from the plan in that, as explained below, we do not perform a direct comparison of the effects in Hypothesis 1 and Hypothesis 2 as ex-post reasoning suggests the two are hardly comparable.

3.2 Experimental design

We run our experiment on a target sample size of 1000 individuals representative of the German population along the age, gender, income and geographic dimensions.[9] We programmed the experiment in Qualtrics and delegated the distribution of the link to the experiment to the panel company Respondi.[10]

3.2.1 Design

The experimental design adopted for the advisees and described here can be visualised on the right hand side of Figure 3.1. We provide a description of the experimental design used to collect the advice in Section 3.2.2. An English transcription of the German survey is available in Appendix 3.5.3.

Phase 1: Demographics At the very beginning of the session, we collect information about the respondents' gender, age, German state of residence and family income, which we use to ensure our sample is representative along these dimensions. We further collect, at the beginning, information about the respondents' family status and household size.

Phase 2: Experimental conditions, receiving advice and lottery choice Our aim is that of identifying how the advisor's monetary status determines the influence of their advice on the (very personal) choice of an advisee. Our design relies on each participant deciding what portion of their endowment (if any) to invest in a lottery yielding triple the amount invested or zero with equal probabilities. The participants keep for sure the fraction of the endowment that was not invested. The lottery is first fully and transparently described.

Next, the respondents receive a randomly selected piece of advice consisting of what fraction of their endowment they should invest according to an advisor. We experimentally vary which out of four types of advisor the advice originates from: One who has a high or a low endowment which was either assigned randomly or based on

[9] We detect a minimum effect of the advisors' status on standardised outcomes (see Section 3.3) of MDE=0.25 (25% of a standard deviation) at power $\pi = 0.8$ and $\alpha = 0.05$.

[10] https://www.respondi.com/EN/

the result of a cognitive ability test. A transcription of the information the advisee received is reported here below:

> "Additional to the participation fee of 90 mingle points, you will be assigned an **additional budget of mingle-points** that you can use later on.
>
> **This additional budget will consist of either 50 or 100 mingle-points.** Whether you will be assigned 50 or 100 mingle points will be decided randomly by the software.
>
> [*Follows a brief description of the lottery's odds and outcomes; see Appendix 3.5.3*]
>
> Before you decide your own lottery investment, we would like to show you an **advice** that has been forwarded by a person who has made their lottery investment in the same lottery at an earlier point of time.
>
> **This person advised you to invest {a percentage is displayed} of your budget.**
>
> This person was assigned a budget of **50 [100] mingle-points**."

The respondents are hence made aware of which type of advisor left the advice they received. Notice further that, as explained below, the size of their endowment is not yet revealed to the advisees when choosing their investment. To obtain sharp predictions and tests, the advice to be passed on to respondent i takes two values $\hat{s}_i \in \{30\%, 70\%\}$, for the investment of relatively low or relatively high proportions of endowment. Each advice originates from advisors in each cell of our earned/random $\times$ high/low endowment advisor type.

This strategy offers three advantages. First, the advice offered is relatively far from the focal investment of 50% of one's endowment. Second, we offer advice which can be palatable both to relatively more or less risk averse individuals. Third, we ensure that the advice used to deliver our experimental conditions is orthogonal to the advisors' (observable or unobservable) characteristics, their cognitive ability in particular.

Next, respondent i, having received advice $\hat{s}_i$ and knowing what type of advisor it comes from, make their investment choice $s_i \in \{0\%, 1\%, ..., 100\%\}$. We measure the impact of advice as the absolute distance $|s_i - \hat{s}_i|$ between participant i's investment choice and the advice received. The lottery is *not yet realised*.

Notice that to avoid design effects, we assign the advisees either a low or a high endowment (same sizes as for the advisors), though the endowment is always randomly assigned. However, the advisees are communicated the size of their endowment only *after* they made their investment to preserve the exogeneity of the investment choice with respect to their own endowment. As all these design features are openly communicated to the advisees, we moreover ensure they don't feel unfairly treated relative to the advisors (they too have a chance at a high or low endowment). We can thus cleanly attribute treatment effects to our experimental variation: The size of the advisor's endowment.

Phase 3: Cognitive Reflection Test After the respondents chose their lottery investment but before the lottery is realised, we administer the Cognitive Reflection Test (CRT) developed by Frederick (2005) to elicit their cognitive ability. The test consists of five simple mathematical questions trading off individuals' ability to provide a reasoned *correct* answer over an intuitive though *incorrect* one. The higher the number of correct answers provided, the greater the respondent's cognitive ability. Frederick (2005) shows performance on the test to correlate strongly with other measures of individuals' cognitive ability such as SAT scores. The respondents were given 5 minutes to answer all questions and were remunerated for each correct answer. Notice that because the CRT test is administered after the lottery, participation in this task cannot influence investment decisions, nor can it be in turn affected by the lottery's outcome as it is not yet revealed to the participants. The CRT is particularly relevant in this framework, as it measures the individuals' ability to reason through the details of a problem in search for the correct solution among easier and immediately available ones.

Phase 4: Lottery realisation, further demographics and debriefing After the respondents participated in the CRT, the lottery is visualised on their screen as a "wheel

of fortune" which they can activate by clicking on a button. This implementation helps the respondents graphically visualise the lottery: The wheel is split in twelve equal fields, half of which are coloured green and read "Triple", and the other half is coloured red and say "Zero". The outcome is determined by the position in which the wheel stops, which is in turn determined by a number randomly extracted by the background software. At the end of the survey we collect information about the respondents' education level and their employment status. Finally, we debriefed the respondents on their assigned endowment, on the outcome of their choices and on their earnings.

3.2.2 Collecting the advice

We collected advice from a small number of respondents ($N_{adv} = 99$) on the 10^{th} of March, 2021, before collecting data from our target group, the advisees. In order to be able to collect advice from the advisors in accordance with our experimental design, the sequence of tasks differs slightly from that given to the advisees (see the left hand side of Figure 3.1).

We randomly allocated the advisors to two exclusive groups: One group would be randomly allocated an endowment, while the second would be assigned an endowment based on their score on the CRT. Different from the advisees, the advisors faced the CRT before choosing their investment and leaving advice.

After having been assigned or earned their endowment but without knowing yet its value, the advisors are given a complete and detailed description of the lottery task (including the outcomes and the odds) and of their choice set. They are then asked to leave advice, in the form of what fraction of the endowment to invest, which might or might not be passed on to a respondent who will face the same investment choice at a later point in time. Leaving the advisor ignorant of the size of their endowment minimises the risk that advice might be sensitive to the advisor's endowment size. We allow the advisors to pick their advice from the set {0% (no investment), 30%, 50%, 70%, 100% (full investment)}, to obtain a manageable advice space. We moreover inform the advisor that the person who might receive their advice might have either a low or a high endowment.

Only after having left their advice, the advisors can choose which fraction of their endowment to invest in the lottery. Notice however that the advisors only serve the purpose of allowing us to construct the experimental manipulations to which the advisees, our population of interest, will be exposed without deception. The advisors' data will hence not be part of any of our analyses.

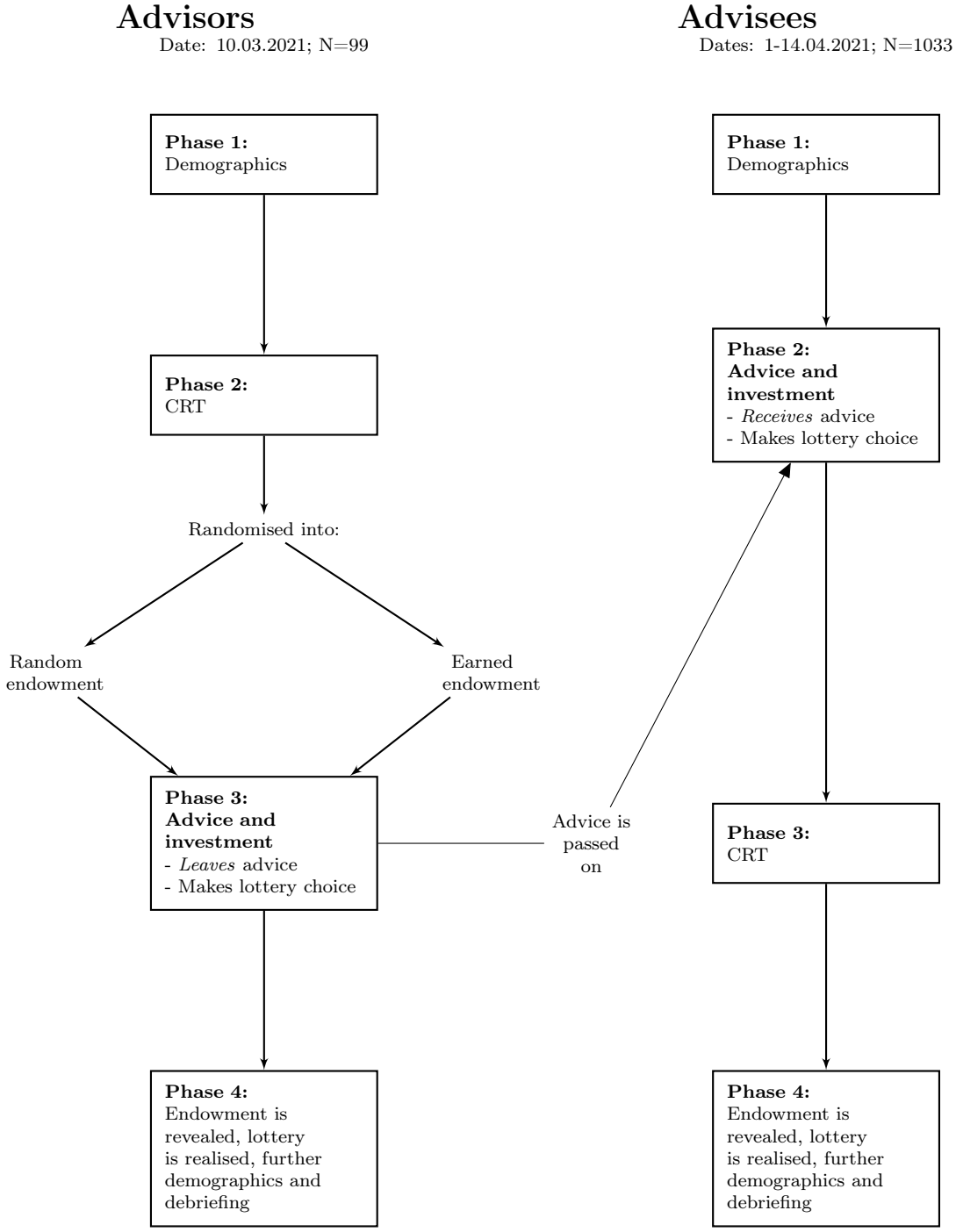


Figure 3.1. Flowchart of the subjects' progress through the experiment

3.3 Empirical strategy and results

We investigate the impact of the advice received in terms of the average proximity of the chosen investment strategy to the one received as advice. Denote with s_i the proportion of endowment invested in the lottery by respondent i, and the proportion received as advice by respondent i with $\hat{s}_i$. Then, our variable of interest can be written as:

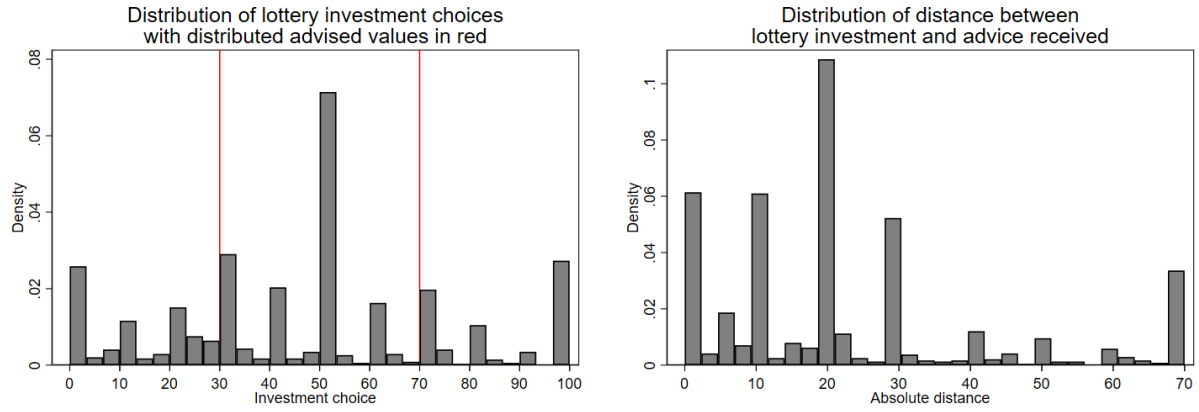
$$y_i = |s_i - \hat{s}_i|.$$

Our main analyses will then rely on the double-bounded Tobit estimation of

$$y_i = \beta_0^r + \beta_1^r HighE + \beta_2^r \hat{s}_i + \beta_3^r CA + \beta^r X + \varepsilon, \tag{3.1}$$

where $r \in \{random, earned\}$ indicates whether the advice received originates from an advisor who was randomly assigned or who earned their endowment and *HighE*, our experimental condition indicator, takes value 1 when the advisor's endowment is high. This way, while we do not observe the influence of advice on individual advisee's choices, β_1^r informs us whether the choices are closer to the advice received for a given advisor's endowment size compared to the other on average.

Here, $0 \leq y_i \leq 70$ is bounded below and above by the minimum (exact matching) and maximum (full or no investment) distance from the advice received, *CA* denotes the respondent's cognitive ability score, and the superscripts r and e indicate whether respondent i received advice from advisors who were, respectively, randomly assigned or who earned their endowment. Finally, X is a vector including the respondent's age, region of residence, equivalent household income, gender and an indicator for tertiary education. Estimating equation (3.1) allows us to test Hypotheses 1 and 2, as summarised in Table 3.1. The same specifications will be used in auxiliary estimations relying on Probit and OLS models to provide additional insights and further support our findings.



(a) Distribution of investments in percentage points from 1 to 100, and in red the distributed advised values.

(b) Distribution of the absolute distance between the investment choice and the advice received.

Figure 3.2. Distribution of investment choices and absolute distance from advice over the entire sample.

Hypothesis:	**Statistical test**
Impact of monetary status	
Hypothesis 1:	$\beta_1^r < 0 \quad \vert r = random$
Hypothesis 2:	$\beta_1^r < 0 \quad \vert r = earned$

Table 3.1. The hypotheses and the corresponding tests

Figure 3.2 depicts the distribution of investment choices (Panel (a)) and of the absolute distance from the advice received (Panel (b)). From Panel (a), we see that the distribution is fairly uniform over the entire investment range, save for evidence of bunching at salient round values, and the obvious focal investment of 50% of the endowment. These observations are reflected in Panel (b). Recall that the advice distributed was in two levels: 30% or 70% of one's endowment. It is obvious thus that among the most frequent absolute distance values are 70 and 30 (distance of advice from extreme choices), 20 (distance of both advices from the salient 50% investment), and 0 (exact matching of advice). The distance value 10 emerges as an interesting behavioural regularity.

In what follows, we will first test whether the type of advisor the advice originates from impacts individuals' choices' placement in these distributions by testing for the impact

of pure monetary status, i.e. when status is assigned randomly, on the advisors' ability to influence others' decision. We will then then turn to the impact of monetary status when it is determined by the advisor's success in the CRT cognitive ability task.

3.3.1 The impact of monetary status

Figure 3.3 displays the average distance between individuals' investment choice and the advice they received from advisors who were randomly allocated their endowment. We disaggregate over the advisors' endowment size. It is already evident from the figure how our respondents are on average 3.701 percentage points closer to the advice received when the latter originates from an advisor who had a high endowment (two-sided t-test p-value=0.034). Because of the bounded nature of the variable and its construction (see Figure 3.2), though seemingly low, this difference is a strong indication of clear behavioural impacts of the advisors' status on individuals' choices in the expected direction.

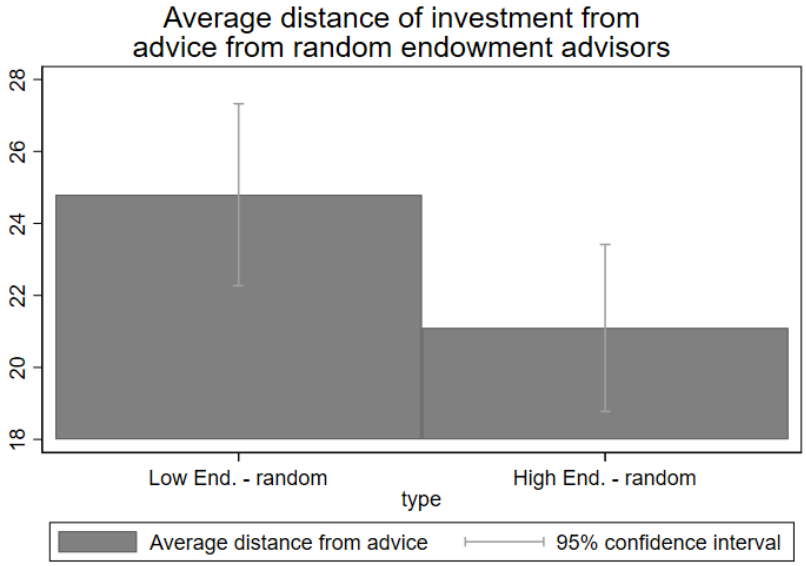


Figure 3.3. Average distance of investment choice from investment advice by type of advisor who were randomly allocated their (high or low) endowment.

We investigate this observation more formally in Table 3.2, reporting the results of double-bounded Tobit estimations of model (3.1). These analyses account for the impact of advice on both the probability of individual i being "bounded" (i.e. perfectly matching

Table 3.2. Effect of advisor's endowment size on advice influence when endowments are randomly assigned

	(1)	(2)	(3)
Distance of individual choice from advice: $y_i = \lvert s_i - \hat{s}_i \rvert$			
Baseline: Low end. advisor			
High end. advisor	-4.633**	-4.559**	-4.306**
	(2.124)	(2.098)	(2.087)
High investment advice		7.117***	7.437***
		(2.098)	(2.128)
Constant	24.475***	20.891***	28.328***
	(1.498)	(1.818)	(4.830)
Marginal effects: $E(y \mid \mathbf{x})$:			
Advisor highly endowed	-3.713**	-3.655**	-3.452**
	(1.695)	(1.675)	(1.667)
Demographics	No	No	Yes
No. of Obs.	516	516	516

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Robust standard errors in parentheses. Tobit regression of absolute difference between advised and actually invested share of endowment on advisor's endowment size, the type of advice received and our individual controls (age, female, equivalent household income tertile, tertiary education, CRT score, German state of residence). Reported marginal effects are computed unconditional on y_i and can be directly interpreted in terms of percentage points difference between investment and advice.

the advice or choosing an investment level at the maximum distance from the advice) and for the proximity to the advice received given that the advisee's choice is not bounded.[11] As we here focus on the impact of the pure monetary status, we include only data from the advisees whose advisors were randomly assigned the endowment.

The results in Table 3.2 demonstrate that the distance between advised and actual invested endowment share is systematically smaller if the advisor was endowed with a randomly assigned high endowment. These findings are robust to the inclusion of controls for the type of advice received (a relatively low or high proportion of one's endowment) and for the respondents' socio-demographic background.[12]

[11] Simple OLS estimations of model (3.1) confirming the results reported in Table 3.2 are reported in Tables 3.5 in Appendix 3.5.1. Notice that difference-in-differences estimations yield the same findings and are available on request.

[12] Appendix 3.5.2 reports the results of further analyses of the probability with which the advisees' choice exactly matches the advice received across treatments. These analyses confirm all the findings presented here.

Worth noting is that advice for investments of relatively larger shares of one's endowment is followed less than advice for smaller investments.

We summarise these findings in Result 1, confirming our first hypothesis that advisors with a high endowment have an influence premium over advisors with a low one when endowments are randomly assigned:

Result 1. *Advice originating from individuals who were randomly assigned a high endowment is followed more closely than that originating from individuals who were randomly assigned a low endowment.*

3.3.2 Earned monetary status

We now discuss the impact of endowment size on the influence of advice originating from advisors who earned their endowment as outcome of a cognitive ability test. The very different nature of differences in endowment size, this time a signal of the individual's underlying cognitive ability, advises to exercise caution in comparing the influence of randomly assigned and earned endowments. The nature of the two are in fact intrinsically different: We can in fact talk about the "impact of status" in the first case only, while we should more properly talk about the "impact of revealed cognitive ability" in the second.

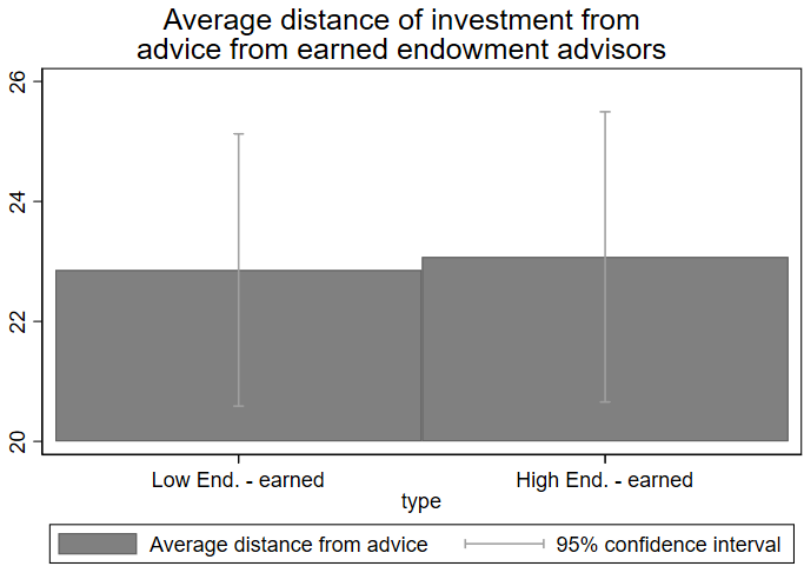


Figure 3.4. Average distance of investment choice from investment advice by advisor's endowment size (high or low endowment), where high (low) endowment corresponds to the high (low) score in the CRT.

Figure 3.4 plots the average distance between individuals' investment choices and the advice received when the advisors earned their endowments. We again disaggregate over the size of the advisor's endowment. Clearly, advice from high or low endowment advisors do not in this case impact investment choices differently (difference=0.218, two-sided t-test p-value=0.897).

Table 3.3 repeats the analysis from Table 3.2, that is the Tobit regression of absolute difference between the investment and the received advice on advisor's endowment size and the set of controls, but now with the data set restricted to cases where the advisors earned their endowment.[13] The estimates show that in this case the impact of advice does not differ according to whether it originated from a high-endowed or low-endowed advisor. Controlling for the type of advice received and for the respondents' socio-demographic background is inconsequential. These findings offer evidence *against* our second hypothesis: high monetary status, this time linked to the advisor's cognitive ability, does *not* grant its possessor greater influence than low status advisors.

[13]Corresponding OLS regressions, which can be found in Table 3.8 in Appendix 3.5.1, confirm the findings here presented.

Thus in contrast to when randomly assigned, monetary status as a signal of the advisors' cognitive ability does not appear to modify the impact of advice.

We hence formulate our second result as follows:

Result 2. *We find no evidence for an influence premium enjoyed by high over low status advisors when status is a signal of their underlying cognitive ability.*

Table 3.3. Effect of advisor's endowment size on advice influence when endowments are assigned based on the CRT score

	(1)	(2)	(3)
	Distance of individual choice from advice: $y_i = \lvert s_i - \hat{s}_i \rvert$		
Baseline: Low end. advisor			
High end. advisor	0.255	0.113	-0.274
	(1.980)	(1.946)	(1.951)
High investment advice		7.620***	6.981***
		(1.946)	(1.957)
Constant	22.010***	18.334***	18.480***
	(1.416)	(1.683)	(4.353)
Marginal effects: $E(y \mid \mathbf{x})$:			
Advisor highly endowed	0.211	0.094	-0.226
	(1.633)	(1.607)	(1.608)
Demographics	No	No	Yes
No. of Obs.	517	517	517

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Robust standard errors in parentheses. Tobit regression of absolute difference between advised and actually invested share of endowment on advisor type indicators, the type of advice received and our individual controls (age, female, equivalent household income tertile, tertiary education, CRT score, German state of residence). Reported marginal effects are computed unconditional on y_i and can be directly interpreted in terms of percentage points difference between investment and advice.

3.3.3 Education and cognitive ability

Our *earned endowment conditions* rely on the distinction between advisors who scored high or low on a cognitive ability test. The purpose of these conditions is that of broadening in a natural direction the perspective within which to read the results obtained from our main experiment (the *random endowment conditions*): Do we observe an influence premium both when status is randomised and when it is instead the expression of the individual's underlying ability? High-status advisors might, in the latter case, enjoy of an influence premium over low status advisors for being seen as

better equipped to provide investment advice. A relevant question to ask is then whether the same principle applies to the advisees. It stands to reason that advisees scoring high on the CRT might respond differently to advice and to the different sources of advice. This mechanism might undermine our efforts in uncovering the relationships of interest by pushing the estimated impact of the advice towards zero.

In what follows, we therefore repeat the analyses reported above by splitting the sample along the CRT performance score (below or equal and above average). Recall that the CRT was administered *after* the subjects made their investment choice, but *before* the outcome of the investment was realised. Therefore, performance on the CRT cannot have influenced the investment, nor can the outcome of the investment have influenced performance on the CRT. We further complement this analysis with a further split based on whether the advisee has at least some years of tertiary education studies or not.[14] Table 3.4 displays the results of these analyses for the case where advisors' endowments were assigned randomly.

Indeed we observe that high CRT scorers do not respond systematically differently to the advice they received from any advisor type. On the other hand, individuals who scored below average in the CRT do respond to advice from advisors who were randomly assigned a high endowment by choosing an investment significantly closer to the advised proportion.

A more nuanced picture emerges if we split the sample according to education level as in columns 3 and 4 of Table 3.4. We recognise the same pattern just discussed: Individuals with at least some tertiary education do not respond differently to advice from either high or low endowed advisor. Conversely, individuals with secondary education or less do alter their reaction to advice according to the type of advisor it originates from: Advisees receiving advice from a high endowed advisors are on average 10.9 percentage points closer to the advice they received than advisees receiving advice from randomly allocated low endowment advisors. Considering the feasible range of y_i, the impact is large and clearly indicates that the group of respondents driving our Result 1 are those with a lower level of education.

[14] Our findings are robust to the choice of cut-off education level.

Table 3.4. Split over the advisees' CRT score and education: Effect of advisor's type on advice influence when endowments are assigned randomly

	(1)	(2)	(3)	(4)
	$y_i = \lvert s_i - \hat{s}_i \rvert$			
	CRT score		Education	
	$\geq$ average	$<$ average	$>$ secondary	$\leq$ secondary
Baseline: Low end. advisor				
High end. advisor	-4.379	-4.497*	1.216	-10.900***
	(3.151)	(2.668)	(2.820)	(2.990)
High investment advice	4.516	10.689***	3.211	11.598***
	(3.203)	(2.719)	(2.825)	(3.052)
Constant	30.844***	28.880***	33.958***	26.524***
	(6.575)	(5.945)	(6.546)	(6.138)
Marginal effects: $E(y \mid \mathbf{x})$:				
Advisor highly endowed	-3.421	-3.689*	0.969	-8.791***
	(2.454)	(2.181)	(2.247)	(2.362)
Demographics	Yes	Yes	Yes	Yes
No. of Obs.	266	250	290	226

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Robust standard errors in parentheses. Split sample analyses along the CRT score (above or below average) and education level (at least some tertiary or at most secondary). Tobit regression of absolute difference between advised and actually invested share of endowment on advisor type indicators, the type of advice received and our individual controls (age, female, equivalent household income tertile, tertiary education, German state of residence). Reported marginal effects are computed unconditional on y_i and can be directly interpreted in terms of percentage points difference between investment and advice.

Repeating the analysis for the case in which the advisors earned their endowments (not reported for brevity) we find results consistent with Table 3.3: We detect no significant impact difference between advice originating from high status advisors who scored high in the CRT test and advice originating from advisors who instead scored low, irrespective of the advisees' CRT score or education level. These findings provide further evidence in support of Result 2.[15]

[15]Irrespective of whether an advisor was assigned or earned their endowment, advice for a relatively high investment is followed significantly less. From Tables 3.3 and Table 3.10, it is clear that this difference is driven by advisees who scored low on the CRT test and/or have a lower education level. This finding resonates well with previous research studying linking cognitive abilities and risk taking. Dohmen et al. (2010) and Benjamin et al. (2013) (among others) find in fact a negative correlation between risk aversion and various measures of cognitive ability.

3.4 Discussion and concluding remarks

Status rankings are ubiquitous in human organizations, as is the evidence that individuals occupying the higher rungs of the social ladder have been accorded privilege, prestige and influence throughout the course of history (for instance, Cheng and Tracy (2013)). Evolutionary theories in psychology and biological science posit that recognition of status differentials and the attribution of influence premia to high status individuals form together part of an evolutionary strategy aimed at enabling and facilitating the transmission of socially valuable traits and behaviours across individuals and generations. These theories take higher status to be accorded to the fittest individuals, similar to what observed among groups of primates and elsewhere in the natural world (Henrich and Gil-White (2001)). This paper addressed the question of how much of the relationship between status and fitness survives among humans. In other words, we asked whether status recognition and the attribution of status privilege (in terms of influence over others' behaviours) are hard-wired into interpersonal relationships to the extent that the existence of an underlying fitness basis for status becomes irrelevant.

We adopted an experimental design allowing us to investigate how much individuals respond to status when it is randomly assigned, and complemented this investigation with an analysis of individuals' responses to status when the latter is instead a signal for the status holder's underlying cognitive ability. We were thus able to measure influence differences between high and low status individuals in a framework in which the potential confounds generated by status beliefs in the real world were completely removed.

Our findings support the hypothesis that high status, even when product of chance, accords its possessor a strong influence premium over others' choices: The respondents in our experiment follow the recommendations of strangers randomly endowed with a high monetary status more than advice from others whose status was randomly assigned to be low. Importantly, we built our experiment in such a way as to render the recommendations received by our respondents completely void of any meaning: No universally valid advice can, in our setup, meaningfully be given to an individual without knowing their risk preferences or endowment. This fact helps us highlight the purely status-based origin of

influence differentials. More in-depth analysis reveals moreover that the regularities here uncovered are driven by the less educated respondents. We moreover uncover that high and low status are statistically indistinguishable, in terms of influence, when linked to the status holders' underlying cognitive ability. We again advice caution in comparing the impact of advice across earned and random conditions due to the very different nature of their signal. Many different mechanisms might come into play once monetary status is grounded in (cognitive) ability, linked among others to spiteful advice rejection due to the implied cognitive ability differentials (though recall that our advisees had not yet taken the cognitive ability test when receiving advice), or to the different origin of the endowment (advisees' endowment was always randomly assigned).

These results shine a new light on the nature of status and of status privileges, suggesting that status recognition and the tendency to accord it privileges might indeed be evolutionarily hard-wired in human psychology. Crucially, despite the adaptive and evolutionary purpose of this mechanism among the first humans as well in current mammalian groups (Smith (1982), Witkower et al. (2020)), the connection between individual fitness and status is made noisy by the accumulation and inheritability of conspicuous status sources ((Hill, 1984), Cheng and Tracy (2013)).The implications of this observations are well borne by our data: Influence, and perhaps deference and other material and non-material advantages, might be accorded to individuals based purely on their perceptible status attributes (e.g. conspicuous consumption) without knowledge, or even in the absence, of underlying socially desirable status-defining attributes or fitness. At the extreme, power relations between individuals could in part be based upon material status differentials and not on differences in characteristics relevant for the relationship itself such as, for instance, competence.

A few previous findings from an incipient literature on the behavioural consequences of status differentials are consistent with and speak towards the findings here reported. For instance, in Martinangeli (2021) and Rockenbach et al. (2021) low status individuals tend to expect greater cooperativeness and to attribute better "qualities" to high status others and to condition their behaviours on (their expectations of) those of the latter, even in minimal experimental settings. We hope our findings will encourage

researchers to broaden the scope of this literature to further our understanding of the origin of the psychological and evolutionary components of status, of its behavioural and psychological consequences, and of its consequences on social and economic relationships (e.g. workplace relations, policy demand and political preferences). A natural next step in the investigation of the patterns uncovered here is to understand which channels express the evolutionary mechanism here outlined: It could for instance be an increased focality of the message originating from high monetary status individuals, or an increase in the perceived authoritativeness of its content, or envy-driven mimicking intentions (e.g., Ronayne and Sgroi (2018)). It is beyond the scope and outside of what feasible with this paper to tease out alternatives, which we leave for the future. Clearly, the evolutionary perspective we started out with and which guided us in the formulation of the hypothesis and in the design of the experiment need not necessarily be the only rationale behind our findings. While we find it appealing and a promising explanation underlying our results, we remain open to the possibility that alternative mechanisms might might be uncovered by future investigations.

3.5 Appendix

3.5.1 OLS regressions

Following four tables report the results of OLS regressions analogous to the Tobit estimates reported in Section 3.3. The patterns discussed there are here confirmed.

Table 3.5. Effect of advisor's type on advice influence when endowments are assigned randomly

	(1)	(2)	(3)
Distance of individual choice from advice: $y_i = \lvert s_i - \hat{s}_i \rvert$			
Advisor highly endowed	-3.702**	-3.654**	-3.482**
	(1.741)	(1.722)	(1.755)
High advice		6.146***	6.421***
		(1.722)	(1.789)
Constant	24.798***	21.726***	28.041***
	(1.231)	(1.491)	(4.064)
Demographics	No	No	Yes
Observations	516	516	516
R-squared	0.009	0.033	0.075

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors in parentheses.
OLS regression of absolute difference between advised and actually invested share of endowment on advisor type indicators, the type of advice received and our individual controls (age, female, equivalent household income tertile, tertiary education, CRT score, German state of residence).

Table 3.6. Effect of advisor's type on advice influence when endowments are assigned based on the CRT score

	(1)	(2)	(3)
	Distance of individual choice from advice: $y_i = \lvert s_i - \hat{s}_i \rvert$		
Advisor highly endowed	0.218	0.067	-0.155
	(1.688)	(1.662)	(1.706)
High advice		6.963***	6.427***
		(1.662)	(1.711)
Constant	22.857***	19.514***	19.917***
	(1.209)	(1.432)	(3.811)
Demographics	No	No	Yes
Observations	517	517	517
R-squared	0.000	0.033	0.068

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors in parentheses.
OLS regression of absolute difference between advised and actually invested share of endowment on advisor type indicators, the type of advice received and our individual controls (age, female, equivalent household income tertile, tertiary education, CRT score, German state of residence).

Table 3.7. Split over CRT score and education: Effect of advisor's type on advice influence when endowments are distributed randomly

	(1)	(2)	(3)	(4)
	$y_i = \lvert s_i - \hat{s}_i \rvert$			
	CRT score		Education	
	$\geq$ average	$<$ average	tertiary	$\leq$ secondary
Advisor highly endowed	-3.515	-3.787	1.260	-9.283***
	(2.599)	(2.402)	(2.383)	(2.649)
High advice	3.882	9.325***	2.683	10.165***
	(2.650)	(2.444)	(2.387)	(2.697)
Constant	30.514***	28.620***	32.718***	26.911***
	(5.389)	(5.374)	(5.482)	(5.454)
Demographics	Yes	Yes	Yes	Yes
Observations	266	250	290	226
R-squared	0.117	0.137	0.105	0.195

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors in parentheses.
Split sample analyses along the CRT score (above or below average) and education level (at least some tertiary or secondary). OLS regression of absolute difference between advised and actually invested share of endowment on advisor type indicators, the type of advice received and our individual controls (age, female, equivalent household income tertile, tertiary education, German state of residence).

Table 3.8. Effect of advisor's type on advice influence when endowments are assigned based on the CRT score

	(1)	(2)	(3)
Distance of individual choice from advice: $y_i = \lvert s_i - \hat{s}_i \rvert$			
Advisor highly endowed	0.218	0.067	-0.155
	(1.688)	(1.662)	(1.706)
High advice		6.963***	6.427***
		(1.662)	(1.711)
Constant	22.857***	19.514***	19.917***
	(1.209)	(1.432)	(3.811)
Demographics	No	No	Yes
Observations	517	517	517
R-squared	0.000	0.033	0.068

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors in parentheses.
OLS regression of absolute difference between advised and actually invested share of endowment on advisor type indicators, the type of advice received and our individual controls (age, female, equivalent household income tertile, tertiary education, CRT score, German state of residence).

3.5.2 Further analyses - exact matching of advice

Random endowments The Tobit analysis reported in Table 3.2 and Table 3.3 accounts for the impact of advice on both the probability of individual i being "bounded" (i.e. perfectly matching the advice or choosing an investment level at the maximum distance from the advice) and for the proximity to the advice received given that he or she is not bounded. In order to investigate whether monetary status of the advisor impacts the probability of exactly matching the advice received, we build an indicator $M_i = \mathbf{1}(y_i = 0)$ for subject i exactly matching the advice received. Table 3.9 shows the results of its Probit regression on our condition indicators and our set of control variables.

Table 3.9. Effect of advisor's endowment on probability of exactly matching the advice when endowments are assigned randomly

	(1)	(2)	(3)
	Matched: $\Pr(M_i = 1)$		
Baseline: Low end. advisor			
High end. advisor	0.264*	0.269*	0.341**
	(0.147)	(0.148)	(0.165)
High investment advice		-0.343**	-0.472***
		(0.149)	(0.169)
Constant	-1.346***	-1.197***	-1.210***
	(0.110)	(0.126)	(0.374)
Marginal effects: $E(y \mid \mathbf{x})$:			
Advisor highly endowed	0.050*	0.050*	0.057**
	(0.028)	(0.027)	(0.027)
Demographics	No	No	Yes
No of Obs.	516	516	484

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors in parentheses.
Probit regression of fully matching the advice received on advisor type indicators, the type of advice received and our individual controls (age, female, equivalent household income tertile, tertiary education, CRT score, German state of residence). Only the data of the advisees whose advisors have been assigned the endowment randomly is included.

When endowments are randomly assigned, we observe from Table 3.9 that advice originating from an advisor whose endowment is high is significantly more likely to be matched exactly than advice from a low-endowment advisor. Controlling for the type of advice received (relatively high or low investments) does not significantly alter our

estimates of interest, and shows again that high-investment advice is strongly and significantly less likely to be matched than low-investment advice.

All these findings are robust to the inclusion of socio-demographic background controls. In these aggregate analyses, we observe no gender, education, income or location effects.

Earned endowments Table 3.10 reports the results from the Probit regression of the indicator variable $M_i = \mathbf{1}(y_i = 0)$ on advisor's endowment size and the set of controls, but now focusing only on the cases where the advice was provided by an advisor who earned their endowment. Consistent with the results of the Tobit regression, advisor's size of endowment does not influence the probability of advisor exactly matching provided advice.

Table 3.10. Effect of advisor's endowment on probability of exactly matching the advice when endowments are assigned based on the CRT score

	(1)	(2)	(3)
	Matched: $\Pr(M_i = 1)$		
Baseline: Low end. advisor			
High end. advisor	0.076	0.090	0.182
	(0.148)	(0.149)	(0.168)
High investment advice		-0.262*	-0.246
		(0.150)	(0.166)
Constant	-1.286***	-1.175***	-1.511***
	(0.108)	(0.124)	(0.403)
Marginal effects: $E(y \mid \mathbf{x})$:			
Advisor highly endowed	0.014	0.016	0.028
	(0.027)	(0.027)	(0.025)
Demographics	No	No	Yes
No. of Obs.	517	517	492

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors in parentheses.
Probit regression of fully matching the advice received on advisor type indicators, the type of advice received and our individual controls (age, female, equivalent household income tertile, tertiary education, CRT score, German state of residence). The regression was run on data of those advisees whose advisors have earned their endowment via CRT test. Only the data of the advisees whose advisors have earned their endowment via CRT test is included.

3.5.3 Survey

Survey transcription

The following pages report an English transcription of the survey. Alternatives randomized between respondents are presented in brackets or parentheses, with eventual commentaries to the randomization in grey.

We are a group of independent researchers at one of the institutes of the Max Planck Society.

This survey should (on average) last for 15-20 minutes.

Please note: Your participation in this study is completely voluntary. We will receive only the anonymized form of your answers. This anonymized data will be saved and used only for scientific purposes.

- ○ Yes, I want to participate in this survey
- ○ No, I don't want to participate in this survey

Are you male or female?

- ○ Female
- ○ Male

How old are you?

- ○ 18-29 Years old
- ○ 30-39 Years old
- ○ 40-49 Years old
- ○ 50-59 Years old
- ○ 60 Years old or older

Please select the federal state and your municipality / your city / your region of residence:

What is your family status?
Welchen Familienstand haben Sie?

- ○ Single (Have never married before / widow(er) / separated / divorced)
- ○ Married / registered partnership / living together with a partner

How many persons live in your household (including yourself):

- ○ Adults _____
- ○ Children (below 18 years old) _____

What was the monthly net income in your household in the last year? By this we mean the total income resulting from the sum of wage, income from self-employment, pension, income from public support, income from renting or leasing, housing subsidies, childcare subsidies, and other sources of income, **after deduction of taxes and social security contributions**.

- ○ < 1000€
- ○ 1000€-1999€
- ○ 2000€-2999€
- ○ 3000€-3999€
- ○ 4000€-4999€
- ○ 5000€-5999€
- ○ 6000€-6999€
- ○ 7000€ or more

Additional to the participation fee of 90 mingle points, you will be assigned an **additional budget of mingle-points** that you can use later on.

This additional budget will consist of **either 50 or 100 mingle-points**. Whether you will be assigned 50 or 100 mingle points will be decided **randomly** by the software.

You can use the additional budget that you have been assigned to participate in a lottery. You can yourself decide what proportion of the additional budget you want to invest in the lottery.

The lottery will either **triple or cancel** the amount of points you invested. The **probability of both outcomes is equal.**
Therefore, you have a 50-50 chance of tripling your investment or losing it.
You will keep for sure the points that you decide not to invest in the lottery.

The outcome will be decided randomly by the software via a wheel of fortune.

Note: at this point the value of the additional budget is still unknown, and can thus be either 50 or 100 mingle-points.
You will be told the value of the budget assigned to you, after you have made the lottery investment decision.

The advised investment proportion was randomized to be high or low and originating from a high or low endowment advisor

Before you decide your own lottery investment, we would like to show you an **advice** that has been forwarded by a person who has made their lottery investment in the same lottery at an earlier point of time.
This person advised you to invest 30% [70%] of your budget.

---------- only displayed to participants in the random endowment advisor condition
This person was assigned a budget of **50 |100| mingle-points**.

Participated on: 10.03.2021

---------- only displayed to participants in the earned endowment advisor condition
This person was assigned a budget of **50 |100| mingle-points** based on their score in a **cognitive ability test**.

Participated on: 10.03.2021

The participant who left the advice for you advises you to invest 30% [70%] of your budget.

---------- only displayed to participants in the random endowment advisor condition
This person was assigned a budget of **50 |100| mingle-points** based on their score in a **cognitive ability test**.
Participated on: 10.03.2021

---------- only displayed to participants in the random endowment advisor condition
This person was assigned a budget of **50 |100| mingle-points** {based on their score in a **cognitive ability test**.
Participated on: 10.03.2021

Please select now the amount that <u>you</u> would like to invest in the lottery.

What proportion of your additional budget would you like to invest in the lottery?

	0 10 20 30 40 50 60 70 80 90 100
Please select what part of your budget you would like to invest; from 0% (nothing at all) up to 100% (everything)	

Before you proceed, we would like to ask you to provide answers to the 5 questions on the next screen.
Please answer all the questions. You will have **5 minutes** to provide the answers.
Any question left without an answer will automatically be considered incorrect.
You will receive two additional points for each correct answer.

A bat and a ball cost $1.10 in total. The bat costs a dollar more than the ball.
How much does the ball cost?
Please enter your answer in cents.

If it takes 2 nurses 2 minutes to measure the blood pressure of 2 patients, how many minutes would 200 nurses need in order to measure the blood pressure of 200 patients?

In a lake, there is a patch of lily pads. Every day, the patch doubles in size.
If it takes 48 days for the patch to cover the entire lake, how long would it take for the patch to cover half of the lake?

Max received both the 15th highest and the 15th lowest mark in the class.
How many students are in the class?
Max hat gleichzeitig die 15. beste und 15. schlechteste Note bekommen.

Simon decided to invest $8,000 in the stock market one day early in 2008. Six months after he invested, on July 17, the stocks he had purchased were down 50%. Fortunately for Simon, from July 17 to October 17, the stocks he had purchased went up 75%.
At this point:

- ○ the value of purchased shares is equal to their value at the time of the purchase.
- ○ the value of purchased shares is higher than their value at the time of the purchase.
- ○ the value of purchased shares is lower than their value at the time of the purchase.

Display earnings – Advisee

In addition to your participation fee of 90 mingle points, you have been randomly assigned a budget of _____ mingle-points.
You invested _____% of this (additional) budget in the lottery, which is _____ mingle-points.

Since you lost (won) in the lottery your earnings from the lottery equal _____ mingle-points.

Therefore, you have earned:

- Participation fee: 90 mingle-points
- The part of the extra budget that you have not invested in the lottery: ____ mingle-points
- Earnings from the quiz: ____ mingle-points
- Earnings from the lottery: ____ mingle-points

In total your earnings equal: ____mingle-points
(your earnings are rounded up to the first integer number of mingle-points)

What is your current employment status?

- ○ Full time employed
- ○ Part time emplyoed
- ○ Self employed or enterpreneur
- ○ I am unemployed and am looking for an employment
- ○ I am a student / in vocational training
- ○ I am retired
- ○ I am currently unemployed and am not looking for an employment

What is your most advanced school degree (in general education or vocational training)?

Bibliography

Alexander, R. D. (1987). *The biology of moral systems*. Transaction Publishers.

Álvarez-Benjumea, A. and F. Winter (2020). The breakdown of antiracist norms: A natural experiment on hate speech after terrorist attacks. *Proceedings of the National Academy of Sciences 117*(37), 22800–22804.

Baik, K. H., S. M. Chowdhury, and A. Ramalingam (2020). The effects of conflict budget on the intensity of conflict: An experimental investigation. *Experimental Economics 23*(1), 240–258.

Ball, S., C. Eckel, P. J. Grossman, and W. Zame (2001). Status in markets. *The Quarterly Journal of Economics 116*(1), 161–188.

Bardsley, N. and R. Sausgruber (2005). Conformity and reciprocity in public good provision. *Journal of Economic Psychology 26*(5), 664–681.

Bazzi, S. and C. Blattman (2014). Economic shocks and conflict: Evidence from commodity prices. *American Economic Journal: Macroeconomics 6*(4), 1–38.

Benjamin, D. J., S. A. Brown, and J. M. Shapiro (2013). Who is 'behavioral'? cognitive ability and anomalous preferences. *Journal of the European Economic Association 11*(6), 1231–1255.

Bicchieri, C., E. Dimant, S. Gaechter, and D. Nosenzo (2020). Observability, Social Proximity, and the Erosion of Norm Compliance. CESifo Working Paper Series 8212, CESifo.

Bicchieri, C. and E. Xiao (2009). Do the right thing: but only if others do so. *Journal of Behavioral Decision Making 22*(2), 191–208.

Bicchieri, C., E. Xiao, and R. Muldoon (2011). Trustworthiness is a social norm, but trusting is not. *Politics, Philosophy & Economics 10*(2), 170–187.

Bignon, V., E. Caroli, and R. Galbiati (2017). Stealing to survive? crime and income shocks in nineteenth century france. *The Economic Journal 127*(599), 19–49.

Blattman, C. and E. Miguel (2010). Civil war. *Journal of Economic Literature 48*(1), 3–57.

Bohlken, A. T. and E. J. Sergenti (2010). Economic growth and ethnic violence: An empirical investigation of hindu—muslim riots in india. *Journal of Peace research 47*(5), 589–600.

Bradley, M. M. and P. J. Lang (1994). Measuring emotion: the self-assessment manikin and the semantic differential. *Journal of behavior therapy and experimental psychiatry 25*(1), 49–59.

Braghieri, L. (2021). Political correctness, social image, and information transmission. Technical report.

Brockner, J., S. Nathanson, A. Friend, J. Harbeck, C. Samuelson, R. Houser, M. H. Bazerman, and J. Z. Rubin (1984). The role of modeling processes in the "knee deep in the big muddy" phenomenon. *Organizational Behavior and Human Performance 33*(1), 77–99.

Bull, R. and N. Rumsey (2012). *The Social Psychology of Facial Appearance*. Springer Science & Business Media.

Burke, M., S. M. Hsiang, and E. Miguel (2015). Climate and Conflict. *Annual Review of Economics 7*(1), 577–617.

Bursztyn, L., G. Egorov, and S. Fiorin (2020). From extreme to mainstream: The erosion of social norms. *American economic review 110*(11), 3522–48.

Bursztyn, L., A. L. González, and D. Yanagizawa-Drott (2018). Misperceived social norms: Female labor force participation in saudi arabia. Technical report, National Bureau of Economic Research.

Cason, T. N., W. A. Masters, and R. M. Sheremeta (2018). Winner-take-all and proportional-prize contests: Theory and experimental results. *Journal of Economic Behavior and Organization*.

Chassang, S. and G. Padro-i Miquel (2009). Economic shocks and civil war. *Quarterly Journal of Political Science 4*(3), 211–228.

Chaudhuri, A., S. Graziano, and P. Maitra (2006). Social Learning and Norms in a Public Goods Experiment with Inter-Generational Advice. *The Review of Economic Studies 73*(2), 357–380.

Chaudoin, S., Z. Peskowitz, and C. Stanton (2017). Beyond zeroes and ones: The intensity and dynamics of civil conflict. *Journal of Conflict Resolution 61*(1), 56–83.

Cheng, J. T. and J. L. Tracy (2013). The Impact of Wealth on Prestige and Dominance Rank Relationships. *Psychological Inquiry 24*(2), 102–108. Publisher: Routledge _eprint: https://doi.org/10.1080/1047840X.2013.792576.

Chevalier, J. A. and D. Mayzlin (2006). The effect of word of mouth on sales: Online book reviews. *Journal of marketing research 43*(3), 345–354.

Choi, J.-K. and S. Bowles (2007). The coevolution of parochial altruism and war. *science 318*(5850), 636–640.

Chowdhury, S. M., J. Y. Jeon, and A. Ramalingam (2018). Property rights and loss aversion in contests. *Economic Inquiry 56*(3), 1492–1511.

Christopher, A. N. and B. R. Schlenker (2000). The impact of perceived material wealth and perceiver personality on first impressions. *Journal of Economic Psychology 21*(1), 1–19.

Cialdini, R. B. (2001). *Influence: Science and Practice*. Allyn and Bacon.

Cikara, M. and J. J. Van Bavel (2014). The neuroscience of intergroup relations: An integrative review. *Perspectives on Psychological Science 9*(3), 245–274.

Cohen, T. R. and C. A. Insko (2008). War and peace: Possible approaches to reducing intergroup conflict. *Perspectives on Psychological Science 3*(2), 87–93.

Cook, K. S. (1975). Expectations, evaluations and equity. *American Sociological Review*, 372–388.

Cornes, R. and R. Hartley (2003). Loss Aversion and the Tullock Paradox. *Keele Economics Research Papers*.

Dal Bó, E. and P. Dal Bó (2011). Workers, warriors, and criminals: social conflict in general equilibrium. *Journal of the European Economic Association 9*(4), 646–677.

Dalmaso, M., G. Pavan, L. Castelli, and G. Galfano (2012). Social status gates social attention in humans. *Biology letters 8*(3), 450–452.

Dancygier, R. and E. N. Saunders (2006). A new electorate? comparing preferences and partisanship between immigrants and natives. *American Journal of Political Science 50*(4), 962–981.

Dell, M. (2015). Trafficking networks and the mexican drug war. *American Economic Review 105*(6), 1738–79.

DellaVigna, S., R. Enikolopov, V. Mironova, M. Petrova, and E. Zhuravskaya (2014). Cross-border media and nationalism: Evidence from serbian radio in croatia. *American Economic Journal: Applied Economics 6*(3), 103–32.

DellaVigna, S., J. A. List, U. Malmendier, and G. Rao (2016). Voting to tell others. *The Review of Economic Studies 84*(1), 143–181.

Derks, B., C. Van Laar, and N. Ellemers (2016). The queen bee phenomenon: Why women leaders distance themselves from junior women. *The Leadership Quarterly 27*(3), 456–469.

Dinas, E., V. Fouka, and A. Schläpfer (2021a). Family history and attitudes toward out-groups: evidence from the european refugee crisis. *The journal of politics 83*(2), 647–661.

Dinas, E., V. Fouka, and A. Schläpfer (2021b). Recognition of collective victimhood and outgroup prejudice. *Public opinion quarterly 85*(2), 517–538.

Dohmen, T., A. Falk, D. Huffman, and U. Sunde (2010). Are risk aversion and impatience related to cognitive ability? *American Economic Review 100*(3), 1238–60.

Duarte, J., S. Siegel, and L. Young (2012). Trust and Credit: The Role of Appearance in Peer-to-peer Lending. *The Review of Financial Studies 25*(8), 2455–2484.

Durham, Y., J. Hirshleifer, and V. L. Smith (1998). Do the rich get richer and the poor poorer? experimental tests of a model of power. *The American Economic Review 88*(4), 970–983.

Egas, M., R. Kats, X. Van Der Sar, E. Reuben, and M. W. Sabelis (2013). Human cooperation by lethal group competition. *Scientific reports 3*(1), 1–4.

Elinder, M., O. Erixson, and D. Waldenström (2018). Inheritance and wealth inequality: Evidence from population registers. *Journal of Public Economics 165*, 17–30.

Ellemers, N., H. Van den Heuvel, D. De Gilder, A. Maass, and A. Bonvini (2004). The underrepresentation of women in science: Differential commitment or the queen bee syndrome? *British Journal of Social Psychology 43*(3), 315–338.

Enikolopov, R., A. Makarin, M. Petrova, and L. Polishchuk (2020). Social image, networks, and protest participation. *Networks, and Protest Participation (April 26, 2020)*.

Falk, A. and J. J. Heckman (2009). Lab Experiments Are a Major Source of Knowledge in the Social Sciences. *Science 326*(5952), 535–538.

Fearon, J. D. (2008). Economic development, insurgency, and civil war. In *Institutions and economic performance*, pp. 292–328. Cambridge, Mass. [u.a.] : Harvard Univ. Press.

Fehr, D. (2018). Is increasing inequality harmful? Experimental evidence. *Games and Economic Behavior 107*, 123–134.

Fehr, E. and U. Fischbacher (2003). The nature of human altruism. *Nature 425*(6960), 785–791.

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

Fiske, S. T., A. J. C. Cuddy, P. Glick, and J. Xu (2002). A model of (often mixed) stereotype content: Competence and warmth respectively follow from perceived status and competition. *Journal of Personality and Social Psychology 82*(6), 878–902.

Fouka, V. (2019). How do immigrants respond to discrimination? the case of germans in the us during world war i. *American Political Science Review 113*(2), 405–422.

Fouka, V., S. Mazumder, and M. Tabellini (2021). From Immigrants to Americans: Race and Assimilation during the Great Migration. *The Review of Economic Studies*.

Fouka, V. and M. Tabellini (2021). Changing in-group boundaries: The effect of immigration on race relations in the united states. *American Political Science Review*, 1–17.

Foulsham, T., J. T. Cheng, J. L. Tracy, J. Henrich, and A. Kingstone (2010). Gaze allocation in a dynamic situation: Effects of social status and speaking. *Cognition 117*(3), 319–331.

Frederick, S. (2005). Cognitive Reflection and Decision Making. *Journal of Economic Perspectives 19*(4), 25–42.

Fried, M. H. (1967). *The evolution of political society: An essay in political anthropology*, Volume 7. Random House.

Gerber, A. S., D. P. Green, and C. W. Larimer (2008). Social pressure and voter turnout: Evidence from a large-scale field experiment. *American political Science review 102*(1), 33–48.

Gino, F. (2008). Do we listen to advice just because we paid for it? The impact of advice cost on its use. *Organizational Behavior and Human Decision Processes 107*(2), 234–245.

Gneezy, U. and J. Potters (1997). An experiment on risk taking and evaluation periods. *The Quarterly Journal of Economics 112*(2), 631–645.

Goerres, A., D. C. Spies, and S. Mayer (2020). Immigrant german election study (imges). GESIS Datenarchiv, Köln. ZA7495 Datenfile Version 1.0.1, https://doi.org/10.4232/1.13544.

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

Greiner, B. and M. V. Levati (2005). Indirect reciprocity in cyclical networks: An experimental study. *Journal of Economic Psychology 26*(5), 711–731.

Hainmueller, J. and D. J. Hopkins (2014). Public attitudes toward immigration. *Annual Review of Political Science 17*(1), 225–249.

Hardy, C. L. and M. Van Vugt (2006). Nice guys finish first: The competitive altruism hypothesis. *Personality and Social Psychology Bulletin 32*(10), 1402–1413.

Heffetz, O. (2018). Are reference points merely lagged beliefs over probabilities? Technical report, National Bureau of Economic Research.

Henrich, J. and F. J. Gil-White (2001). The evolution of prestige: freely conferred deference as a mechanism for enhancing the benefits of cultural transmission. *Evolution and Human Behavior 22*(3), 165–196.

Herbst, L., K. A. Konrad, and F. Morath (2015). Endogenous group formation in experimental contests. *European Economic Review 74*, 163–189.

Hidalgo, F. D., S. Naidu, S. Nichter, and N. Richardson (2010). Economic determinants of land invasions. *The Review of Economics and Statistics 92*(3), 505–523.

Hill, J. (1984). Prestige and reproductive success in man. *Ethology and Sociobiology 5*(2), 77–95.

Hirshleifer, J. (1988). The analytics of continuing conflict. *Synthese 76*(2), 201–233.

Hirshleifer, J. (1989). Conflict and rent-seeking success functions: Ratio vs. difference models of relative success. *Public choice 63*(2), 101–112.

Janus, A. L. (2010). The influence of social desirability pressures on expressed immigration attitudes. *Social Science Quarterly 91*(4), 928–946.

Jost, J. T. (2019). A quarter century of system justification theory: Questions, answers, criticisms, and societal applications. *British Journal of Social Psychology 58*(2), 263–314.

Jost, J. T., B. W. Pelham, O. Sheldon, and B. Ni Sullivan (2003). Social inequality and the reduction of ideological dissonance on behalf of the system: Evidence of enhanced system justification among the disadvantaged. *European journal of social psychology 33*(1), 13–36.

Just, A. and C. J. Anderson (2015). Dual allegiances? immigrants' attitudes toward immigration. *The Journal of Politics 77*(1), 188–201.

Kashima, Y., S. Wilson, D. Lusher, L. J. Pearson, and C. Pearson (2013). The acquisition of perceived descriptive norms as social category learning in social networks. *Social Networks 35*(4), 711–719.

Keller, E. and B. Fay (2012). Word-of-mouth advocacy. *Journal of Advertising Research 52*(4), 459–464.

Kim, N. K. (2016). Revisiting Economic Shocks and Coups. *Journal of Conflict Resolution 60*(1), 3–31.

Kimbrough, E. O., K. Laughren, and R. Sheremeta (2020). War and conflict in economics: Theories, applications, and recent trends. *Journal of Economic Behavior & Organization 178*, 998–1013.

Kimbrough, E. O., R. M. Sheremeta, and T. W. Shields (2014). When parity promotes peace: Resolving conflict between asymmetric agents. *Journal of Economic Behavior & Organization 99*, 96–108.

Kong, X. (2008). Loss Aversion and Rent-Seeking: An Experimental Study. *Discussion Papers*.

Konrad, K. A. and F. Morath (2019). Escalation in conflict games: on beliefs and selection. *Experimental Economics*, 1–38.

Koski, J. E., H. Xie, and I. R. Olson (2015). Understanding social hierarchies: The neural and psychological foundations of status perception. *Social Neuroscience 10*(5), 527–550.

Kőszegi, B. and M. Rabin (2006). A model of reference-dependent preferences. *The Quarterly Journal of Economics 121*(4), 1133–1165.

Kőszegi, B. and M. Rabin (2007). Reference-dependent risk attitudes. *American Economic Review 97*(4), 1047–1073.

Krupka, E. and R. A. Weber (2009). The focusing and informational effects of norms on pro-social behavior. *Journal of Economic psychology 30*(3), 307–320.

Kuo, A., N. Malhotra, and C. H. Mo (2017). Social exclusion and political identity: The case of asian american partisanship. *The Journal of Politics 79*(1), 17–32.

Kuran, T. (1997). *Private truths, public lies: The social consequences of preference falsification*. Harvard University Press.

Kuziemko, I., R. W. Buell, T. Reich, and M. I. Norton (2014). "Last-Place Aversion": Evidence and Redistributive Implications. *The Quarterly Journal of Economics 129*(1), 105–149.

Kwan, L. Y.-Y., S. Yap, and C.-y. Chiu (2015). Mere exposure affects perceived descriptive norms: Implications for personal preferences and trust. *Organizational Behavior and Human Decision Processes 129*, 48–58.

Lacomba, J. A., F. Lagos, E. Reuben, and F. van Winden (2014). On the escalation and de-escalation of conflict. *Games and Economic Behavior 86*, 40–57.

Little, A. C. and S. C. Roberts (2012). Evolution, Appearance, and Occupational Success. *Evolutionary Psychology 10*(5).

Maner, J. K., C. N. DeWall, and M. T. Gailliot (2008). Selective Attention to Signs of Success: Social Dominance and Early Stage Interpersonal Perception. *Personality and Social Psychology Bulletin 34*(4), 488–501.

Markussen, T., S. Sharma, S. Singhal, and F. Tarp (2021). Inequality, institutions and cooperation. *European Economic Review 138*, 103842.

Martinangeli, A. F. M. (2021). Do What (You Think) the Rich Will Do: Inequality and Belief Heterogeneity in Public Good Provision. *Journal of Economic Psychology 83*, 102364.

Martinangeli, A. F. M. and P. Martinsson (2020). We, the rich: Inequality, identity and cooperation. *Journal of Economic Behavior & Organization 178*, 249–266.

Martinangeli, A. F. M. and L. Windsteiger (2020). Last word not yet spoken: a reinvestigation of last place aversion with aversion to rank reversals. *Experimental Economics*.

Mattan, B. D., J. T. Kubota, and J. Cloutier (2017). How Social Status Shapes Person Perception and Evaluation: A Social Neuroscience Perspective. *Perspectives on Psychological Science 12*(3), 468–507. Publisher: SAGE Publications Inc.

McBride, M. and S. Skaperdas (2014). Conflict, settlement, and the shadow of the future. *Journal of Economic Behavior & Organization 105*, 75–89.

Meeusen, C., K. Abts, and B. Meuleman (2019). Between solidarity and competitive threat?: The ambivalence of anti-immigrant attitudes among ethnic minorities. *International Journal of Intercultural Relations 71*, 1–13.

Miguel, E. (2005). Poverty and witch killing. *The Review of Economic Studies 72*(4), 1153–1172.

Miguel, E., S. Satyanath, and E. Sergenti (2004). Economic shocks and civil conflict: An instrumental variables approach. *Journal of Political Economy 112*(4), 725–753.

Morgan, J., H. Orzen, and M. Sefton (2012). Endogenous entry in contests. *Economic Theory 51*(2), 435–463.

Morris, S. (2001). Political correctness. *Journal of political Economy 109*(2), 231–265.

Mujcic, R. and A. Leibbrandt (2018). Indirect reciprocity and prosocial behaviour: evidence from a natural field experiment. *The Economic Journal 128*(611), 1683–1699.

Nelissen, R. M. A. and M. H. C. Meijers (2011). Social benefits of luxury brands as costly signals of wealth and status. *Evolution and Human Behavior 32*(5), 343–355.

Neuhold, C. (2020). Wer erbt die blauen serben? https://www.profil.at/oesterreich/wer-erbt-die-blauen-serben/400961525.

Norton, M. I., S. R. Sommers, E. P. Apfelbaum, N. Pura, and D. Ariely (2006). Color blindness and interracial interaction: Playing the political correctness game. *Psychological Science 17*(11), 949–953.

Nowak, M. A. (2006). Five rules for the evolution of cooperation. *science 314*(5805), 1560–1563.

Nowak, M. A. and K. Sigmund (2005). Evolution of indirect reciprocity. *Nature 437*(7063), 1291–1298.

Ohlsson, H., J. Roine, and D. Waldenström (2020). Inherited Wealth over the Path of Development: Sweden, 1810–2016. *Journal of the European Economic Association 18*(3), 1123–1157.

Oulton, N. (1976). Inheritance and the Distribution of Wealth. *Oxford Economic Papers 28*(1), 86–101. Publisher: Oxford University Press.

Papaioannou, K. J. (2017). "hunger makes a thief of any man": Poverty and crime in british colonial asia. *European Review of Economic History 21*(1), 1–28.

Perez-Truglia, R. and G. Cruces (2017). Partisan interactions: Evidence from a field experiment in the united states. *Journal of Political Economy 125*(4), 1208–1243.

Piketty, T. (2011). On the Long-Run Evolution of Inheritance: France 1820–2050 *. *The Quarterly Journal of Economics 126*(3), 1071–1131.

Price, C. R. and R. M. Sheremeta (2011). Endowment effects in contests. *Economics Letters 111*(3), 217–219.

Price, C. R. and R. M. Sheremeta (2015). Endowment Origin, Demographic Effects, and Individual Preferences in Contests. *Journal of Economics and Management Strategy 24*(3), 597–619.

Ridgeway, C. (1991). The social construction of status value: Gender and other nominal characteristics. *Social Forces 70*(2), 367–386.

Ridgeway, C. L. and H. A. Walker (1995). Status structures. *Sociological perspectives on social psychology 281*, 310.

Rockenbach, B., S. Tonke, and A. R. Weiss (2021). Self-serving behavior of the rich causes contagion effects among the poor. *Journal of Economic Behavior & Organization 183*, 289–300.

Ronayne, D. and D. Sgroi (2018). Ignoring good advice. *Warwick Economic Research Papers 1150*.

Schotter, A. (2003). Decision making with naive advice. *American Economic Review 93*(2), 196–201.

Schotter, A. and B. Sopher (2006). Trust and trustworthiness in games: An experimental study of intergenerational advice. *Experimental Economics 9*(2), 123–145.

Schotter, A. and B. Sopher (2007). Advice and behavior in intergenerational ultimatum games: An experimental approach. *Games and Economic Behavior 58*(2), 365–393.

Shariff, A. F., J. L. Tracy, and J. L. Markusoff (2012). (Implicitly) Judging a Book by Its Cover: The Power of Pride and Shame Expressions in Shaping Judgments of Social Status. *Personality and Social Psychology Bulletin 38*(9), 1178–1193.

Shayo, M. and A. Zussman (2011). Judicial Ingroup Bias in the Shadow of Terrorism *. *The Quarterly Journal of Economics 126*(3), 1447–1484.

Sheremeta, R. M. (2010). Experimental comparison of multi-stage and one-stage contests. *Games and Economic Behavior 68*(2), 731–747.

Sheremeta, R. M. (2011). Contest design: An experimental investigation. *Economic Inquiry 49*(2), 573–590.

Sheremeta, R. M. (2013). Overbidding and heterogeneous behavior in contest experiments. *Journal of Economic Surveys 27*(3), 491–514.

Sheremeta, R. M. (2015). Impulsive Behavior in Competition: Testing Theories of Overbidding in Rent-Seeking Contests. *SSRN Electronic Journal*.

Shupp, R., R. M. Sheremeta, D. Schmidt, and J. Walker (2013). Resource allocation contests: Experimental evidence. *Journal of Economic Psychology 39*, 257–267.

Skaperdas, S. (1992). Cooperation, Conflict, and Power in the Absence of Property Rights. *American Economic Review 82*(4), 720–739.

Smith, J. M. (1982). *Evolution now: A century after Darwin*. Freeman.

Stewart, P. A. and J. C. Moore Jr (1992). Wage disparities and performance expectations. *Social Psychology Quarterly*, 78–85.

Strijbis, O. and J. Polavieja (2018). Immigrants against immigration: Competition, identity and immigrants' vote on free movement in switzerland. *Electoral Studies 56*, 150–157.

Tajfel, H., J. C. Turner, W. G. Austin, and S. Worchel (1979). An integrative theory of intergroup conflict. *Organizational identity: A reader 56*, 65.

Tingley, D. H. (2011). The dark side of the future: An experimental test of commitment problems in bargaining. *International Studies Quarterly 55*(2), 521–544.

Toplak, M. E., R. F. West, and K. E. Stanovich (2014). Assessing miserly information processing: An expansion of the Cognitive Reflection Test. *Thinking & Reasoning 20*(2), 147–168.

Tullock, G. (1980). E¢ efficient rent seeking. *Toward a Theory of the Rent-Seeking Society, edited by JM Buchanan, RD Tollison, and G. Tullock, Texas A&M University Press, College Station*.

Valentim, V. (2022). Social norms and preference falsification in a democracy.

Van der Vegt, G. S., J. S. Bunderson, and A. Oosterhof (2006). Expertness diversity and interpersonal helping in teams: Why those who need the most help end up getting the least. *Academy of Management Journal 49*(5), 877–893.

Van der Zwan, R., P. Bles, and M. Lubbers (2017). Perceived migrant threat among migrants in europe. *European Sociological Review 33*(4), 518–533.

Walker, L. S. (2015). Social influence. In *The Blackwell Encyclopedia of Sociology*. G. Ritzer (Ed.).

Wärneryd, K. (2014). *The economics of conflict: Theory and empirical evidence*. MIT Press.

Weber, M. (2001). *The Protestant Ethic and the Spirit of Capitalism*. London: Routledge.

Werb, D., G. Rowell, G. Guyatt, T. Kerr, J. Montaner, and E. Wood (2011). Effect of drug law enforcement on drug market violence: A systematic review. *International Journal of Drug Policy 22*(2), 87–94.

Witkower, Z., J. L. Tracy, J. T. Cheng, and J. Henrich (2020). Two Signals of Social Rank: Prestige and Dominance Are Associated With Distinct Nonverbal Displays. *Journal of Personality and Social Psychology 118*(1), 32.

Wooldridge, A. (2021). *The Aristocracy of Talent*. Penguin UK.

Xiao, E. and C. Bicchieri (2010). When equality trumps reciprocity. *Journal of Economic Psychology 31*(3), 456–470.

Yanagizawa, D. (2009). Propaganda and con ict: Theory and evidence from the rwandan genocide.

Eidesstattliche Versicherung

Ich versichere hiermit eidesstattlich, dass ich die vorliegende Arbeit selbständig und ohne fremde Hilfe verfasst habe. Die aus fremden Quellen direkt oder indirekt übernommenen Gedanken sowie mir gegebene Anregungen sind als solche kenntlich gemacht.

Die Arbeit wurde bisher keiner anderen Prüfungsbehörde vorgelegt und auch noch nicht veröffentlicht. Sofern ein Teil der Arbeit aus bereits veröffentlichten Papers besteht, habe ich dies ausdrücklich angegeben.

Biljana Meiske

München, den 14 März 2022