Social Preferences Across Subject Pools: Students vs. General Population *

Thomas Epper ⓡ Julien Senn ⓡ Ernst Fehr

January 10, 2024

Abstract

The empirical evidence on the existence of social preferences—or lack thereof—is predominantly based on student samples. Yet, knowledge about whether these findings can be extended to the general population is still scarce. In this paper, we compare the distribution of social preferences in a student and in a representative general population sample. Using descriptive analysis and a rigorous clustering approach, we show that the distribution of the general population's social preferences fundamentally differs from the students' distribution. In the general population, three types emerge: an inequality averse, an altruistic, and a selfish type. In contrast, only the altruistic and the selfish types emerge in the student population. We show that differences in age and education are likely to explain these results. Younger and more educated individuals—which typically characterize students—not only tend to have lower degrees of other-regardingness but this reduction in other-regardingness radically reduces the share of inequality aversion among students. Differences in income, however, do not seem to affect social preferences. We corroborate our findings by examining nine further data sets that lead to a similar conclusion: students are far less inequality averse than the general population. These findings are important in view of the fact that almost all applications of social preference ideas involve the general population.

Key Words: Social Preferences, Altruism, Inequality Aversion, Preference Heterogeneity, Subject pools, Sample Selection

JEL Codes: C80, C90, D30, D63

*The "ⓡ" symbol indicates that the authors' names are in certified random order. Thomas Epper greatfully acknowledges the financial support received from the Métropole Européenne de Lille (MEL), which significantly contributed to the conduct and completion of this research. Epper: IESEG School of Management, Univ. Lille, CNRS, UMR 9221 - LEM - Lille Economie Management F-59000 Lille, France (thomas.epper@cnrs.fr). Fehr: Department of Economics, Zurich University. Blümlisalpstrasse 10, 8006 Zurich, Switzerland (ernst.fehr@econ.uzh.ch) Senn: Department of Economics, Zurich University. Blümlisalpstrasse 10, 8006 Zurich, Switzerland (julien.senn@econ.uzh.ch)

1 Introduction

Models of other-regarding behavior (see e.g. Charness and Rabin, 2002; Fehr and Schmidt, 1999; Bolton and Ockenfels, 2000) have received considerable attention in the last decades. However, the empirical evidence on the existence of such other-regarding subjects—or lack thereof—is predominantly based on student samples.[1] But do students really reflect the general population, or are there important differences that should prevent one from generalizing findings from students to the general population? Quite surprisingly, knowledge about the extent and the ways in which the student population and the general population differ in their other-regarding preferences is still scarce.

In this paper, we compare the distribution of other-regarding preferences in a general population sample that is representative of the swiss population, and a sample of students from the largest Swiss university. In both samples, we elicit other-regarding preferences using a large set of incentivized choice situations in which the decision maker has to decide on how to allocate money between herself and an anonymous partner. When comparing student samples with general population samples, the literature has almost exclusively relied on dictator games, i.e. choice situations in which subjects can sacrifice resources to *increase* the payoff of others, to identify social preferences. This design choice is not inconsequential: while such dictator games are well suited to identify altruism and the extent to which individuals are willing to trade-off equality and efficiency (see e.g. Fisman et al., 2007, 2017), they do *not* allow to identify a broad range of other social preferences. For example, inequality aversion (Fehr and Schmidt, 1999; Charness and Rabin, 2002; Bolton and Ockenfels, 2000) implies that individuals might not only be willing to sacrifice some of their own payoff to increase the payoff of those worse off (aversion to advantageous inequality), but also to decrease the payoff of those who are better off (aversion to disadvantageous inequality). Similarly, envious and spiteful individuals are willing to pay to destroy the payoff of others. In order to be able to also identify these classes of preferences, our experimental paradigm also includes several decision situations where the decision maker can pay in order to decrease the payoff of others.

We explore the behavioral heterogeneity in our two samples in two steps. First, we analyze subjects' choices at the descriptive level by inspecting their willingness to pay to increase the payoff of others and their willingness to pay to decrease the payoff of others.

[1] For example, out of the 24 papers studying social preferences that were published in the top five journals in Economics between 2000 and 2010, "*only four report from experiments on non-student samples, and only two of the papers report from experiments performed outside the lab*" (Cappelen et al., 2015, , p.1307).

This descriptive exercise already reveals stark differences between the two samples.

In the general population sample, the majority of the individuals are located in one of three behavioral agglomerations. The first behavioral agglomeration comprises subjects who predominantly equalize payoffs. These individuals are willing to pay *both to increase* the payoff of those worse off *and to decrease* the payoff of those better off for the sake of greater equality. The second behavioral agglomeration comprises subjects who are willing to sacrifice resources to increase the payoff of those worse off, but are *not* willing to pay to reduce the payoff of those better off. Finally, the third behavioral agglomeration comprises individuals who predominantly maximize their own payoff.

These behavioral patterns contrast sharply with the student sample, where *only two* behavioral agglomerations emerge: one comprising students who predominantly maximize their own payoff, and another comprising students who are willing to sacrifice resources to increase the payoff of those worse off but unwilling to pay to reduce the payoff of those better off. Strikingly, the descriptive analysis reveals that almost no student predominantly equalizes payoffs, i.e. a large behavioral agglomeration documented in the general population sample is almost entirely absent from the student sample.

In a second step, we apply a relatively novel nonparametric Bayesian clustering algorithm—the Dirichlet Process means—in order to more rigorously identify preference heterogeneity in the two samples. This analysis largely confirms the results from the descriptive analysis: In the general population, the DP-means algorithm identifies *three* distinct and empirically relevant types with a clear behavioral interpretation: a majority (50.8%) of inequality averse individuals who predominantly equalize payoffs, a smaller share (34.4%) of altruistic subjects who are willing to pay to increase the payoff of those worse off but unwilling to decrease the payoff of those better off, and a minority (14.8%) of subjects who make predominantly selfish choices. In contrast, the clustering algorithm identifies only *two* types in the student sample: a large cluster of predominantly selfish subjects (54.55%), and another type comprising altruistic subjects. That is, in line with the descriptive evidence, the clustering analysis allows us to document that students' social preferences fundamentally differ from those of the general population in the sense that a whole preference type—the inequality averse type—is basically absent in students but very widespread in the general population.

To shed light on the potential drivers of these behavioral differences, we structurally estimate the parameters of a model of other-regarding preferences and we explore the association between these parameters and different socio-demographics that typically distinguish students from the general population. Specifically, we focus on the role of age, income and education—three dimensions that could affect choices in our money allocation task and along which students and the general population widely differ.

This analysis reveals that age is a particularly strong predictor of aversion to advantageous and disadvantageous inequality: younger individuals tend to display lower degrees of other-regardingness, but as individuals grow older, they tend to become more averse to both advantageous and disadvantageous inequality. In addition, a higher education among young individuals (i.e., being a student) further decreases other-regardingness such that aversion against disadvantageous inequality is much lower among students.

The very low level of inequality aversion in our student sample is remarkable and raises the question of how generalizable our findings are. To address this question, we analyzed data from nine additional data sets, collected from eight different published papers that elicited social preferences with the Equality-Equivalence Test (Kerschbamer, 2015)—a test that can cleanly separate inequality aversion from other forms of social preferences. Five of the nine different data sets are based on student samples and four are based on nationally representative samples from Denmark, Germany and the US.

We calculated the population shares of different social preference types in each of these data sets, and the results are remarkably similar to our own findings. Despite the fact that the Equality-Equivalence Test differs from our approach in many ways, this analysis also shows that the share of inequality averse subjects is much lower in students than in the general population. Indeed, our calculations indicate the existence of only a small share of 7-12 percent inequality averse subjects across the five different student samples, while it is much larger in the gerenal population samples, varying from 23 percent in Denmark to 42 percent in the US and roughly 65 percent in Germany. Our analysis of these additional data sets suggests that students are indeed much less inequality averse than the general population.

Overall, these findings highlight that students and the general population fundamentally differ in terms of their social preferences. While previous studies have suggested that students might simply offer a lower bound on the degree of other-regardingness in the general population (see e.g. Falk et al., 2013; Anderson et al., 2013; Bellemare et al., 2011, 2008), our results suggest that relying on student samples can have much bigger implications, in the sense that the share of inequality aversion is vastly underestimated if one generalizes the results from student samples to the general population. Thus, relying on student samples to make predictions about behaviors in the general population in contexts where social preferences are believed to play a role such as, e.g., labor market relations (see e.g. Falk et al., 2006) or voting behavior (see e.g. Tyran and Sausgruber, 2006; Durante et al., 2014), can lead to wrong inferences.

Our study is connected to the literature that compares other-regarding behavior across subject pools. Previous papers have compared the behavior of students and non-students in the trust game (Fehr and List, 2004; Bellemare and Kröger, 2007; Falk et al., 2013; Cappelen et al., 2015), in bargaining and dictator games (Carpenter et al., 2005; Güth et al., 2007; Carpenter et al., 2008; Bellemare et al., 2008, 2011; Cappelen et al., 2015; Fisman et al., 2015; Snowberg and Yariv, 2021) and in social dilemmas (Gächter et al., 2004; Carpenter and Seki, 2011; Stoop et al., 2012; Anderson et al., 2013; Belot et al., 2015; Snowberg and Yariv, 2021).[2,3] However, bargaining games, trust games as well as other social dilemma games do not provide a clean identification of the different types of social preferences because subjects' behavior is affected by beliefs and strategic considerations. Therefore, these games do not allow to make inferences about subject pool differences in altruism or inequality aversion.

We contribute to this literature and extend it in several ways. First, in our design each individual faces several negatively sloped budget lines *and* several positively sloped budget lines in the space of "own-other's payoff". The different negatively sloped budget lines imply different costs of increasing the other player's payoff while different positively sloped budget lines imply varying costs of decreasing the other's payoff. This design feature enables the identification of a much broader range of other-regarding preferences compared to other studies that examined preferences in students and in the general population. For example, in Fisman et al. (2015) individuals only face negatively sloped budget lines which enables identification of different degrees of altruism but envy, spite and aversion to disadvantageous inequality—which all involve a positive willingness to incur cost to decrease others' payoffs—cannot be identified with these budget lines. Likewise, in Bellemare et al. (2008, 2011) proposers in the ultimatum game face only one negatively sloped budget line (with slope -1) which does not enable the identification of the proposers' willingness to pay to reduce the responders' payoff.[4]

[2] Relatedly, Krawczyk (2011) and Abeler and Nosenzo (2015) investigate how monetary incentives and pro-social motives (e.g. appealing to subjects' wilingness to help the researcher) affect participation in experiments.

[3] Fréchette (2016) reviews experiments using non-standard subjects pools, including experiments conducted with animals and individuals living in token economies. See also Camerer (2011) and Engel (2011).

[4] While it is not possible to identify proposers' envy, spite or aversion against disadvantageous

Second, while existing studies have explored differences between students and the general population using either descriptive evidence or structural estimations, our study investigates this question using a non-parametric clustering analysis—the Dirichlet Process Means (DP-means) algorithm. The advantage of this approach is that it requires no ex-ante assumptions on the existing preference types. Instead, the DP-means algorithm endogenously determines the number of preference clusters and assigns each individual to one of the emerging clusters. Thus, this approach enables us to identify the distribution of individuals over preference types with minimal assumptions. It allows us to document, in particular, that the distribution of students' social preferences fundamentally differs from those in the general population in the sense that the share of inequality averse individuals is vastly underestimated when generalizing results from student samples to the general population.

Finally, we also contribute to the literature by shedding light on the drivers of behavioral differences across the subjects pools by showing that a young age and a high education—which typically characterize students—reduce individuals' other-regardingness to such a degree that the inequality averse type completely vanishes among the students in our sample. While this result is admittedly somewhat surprising, the analysis of nine additional data sets suggests that students are indeed much less inequality averse than the general population.

The remainder of the paper is organized as follows: In the next section, we provide details on the experimental design, the samples and the study implementation. We present our results in Section 3. We discuss the role of age, income and education for our results in Section 4 and we provide further discussion in Section 5. We conclude in Section 6.

---

inequality one may be able to identify the existence of proposers' altruism in the ultimatum game if proposers make higher offers than what is required to induce responders to accept the offer (i.e., if proposers overpay responders). However, with only one negatively sloped budget line precise identification may still be difficult. Likewise, while it is possible to identify the responders' willingness to pay to decrease the proposer's payoff (by examining responders' rejection behavior for positive offers), it is not possible to identify the responders' willingness to pay to increase the proposer's payoff. This is due to the fact that responders only face positively sloped budget lines. For each given offer they face one positively sloped budget line with two feasible allocations – the offered allocation or the allocation (0, 0).

2 Experimental design

2.1 Measuring distributional preferences

We elicited respondents' distributional preferences using a series of twelve incentivized money allocation tasks in which participants had to decide how to allocate experimental currency units (ECUs) between themselves and an anonymous other participant of the study. These twelve choice situations systematically vary the cost and the efficiency consequences of redistribution, thereby allowing us to identify a wide range of other-regarding behaviors. Negatively sloped budget lines (where subjects can sacrifice resources to *increase* the payoff of the other participant) allow us to identify behaviors such as altruism and aversion to advantageous inequality. Positively sloped budget lines (where subjects can pay to *decrease* the payoff of the other) allow us to identify behaviors such as envy, spite and aversion to disadvantageous inequality, among others. We provide details on the various budget lines for which subjects had to make a decision in the Appendix A.1. Choice situations were presented to subjects in random order directly on subjects' screens. In each choice situation, subjects could choose between seven interpersonal allocations. They were presented in a way that makes the distributional consequences of each allocation transparent.[5]

2.2 Samples

We collected data from two separate samples.

Broad population sample The broad population sample consists of 815 individuals and is representative of the swiss population with respect to age, gender and geographical area (see Table B.1 in Appendix).[6] Data collection was completed in collaboration with the LINK Institute, a leading company for high-quality market research in Switzerland, in March and April 2017. Subjects in this sample completed the experiment online. They were encouraged to take the study from a place where they would not be distracted, as is common practice.

[5]For an example choice sitution, see Appendix A.2.

[6]More specifically, our broad population sample comprises swiss citizens older than 18 living living either in the French- or the German-speaking part of Switzerland (who make up more than 90% of the Swiss population). Our data does not comprise observations from the Italian and the Rumanisch-speaking parts of Switzerland.

Students sample The student sample comprises 65 students from the University of Zurich and the Swiss Federal Institute of Technology in Zurich. These participants were recruited using h-Root (Bock et al. 2014) and took part to the experiment in December 2018. Descriptive statistics on the student sample can be found in Table B.2 in Appendix. The experiment was administered using the same online platform as for the broad population sample and was completed in the laboratory.

2.3 Implementation

The experiment was computerized and all the instructions were displayed directly on participants' screens. Participants in both samples had to complete the experiment on the same online platform. We incentivized respondents' choices in the money allocation task by implementing one of their decisions at random. In addition, respondents were paid a show-up fee for completing the study until the end. The study lasted about one hour, for which participants in the broad population sample earned an average of CHF 26 while participants in the student sample earned an average of CHF 23. The exchange rate between points and swiss francs was slightly lower in the student sample (100 points = CHF 1.5) than in the general population sample (100 points = CHF 2.5) in order to account for the fact that subjects in the two samples might have different opportunity costs. We discuss why we believe that this and other minor differences in procedures are very unlikely to drive our results in Section 5. Note also that participants in both samples have taken studies in the past. Thus, we can rule out that different past experiences with completing studies might drive any behavioral difference between our samples.

3 Results

We compare the distribution of other regarding preferences in our general population sample with the distribution in our student sample. We proceed in two steps. First, we inspect participants' choices at the descriptive level. This analysis might already provide preliminary insights regarding how the distribution of other-regearding preferences differs in the two samples. Next, we apply a non-parametric Bayesian clustering algorithm, the Dirichlet Process Means, to uncover behavioral heterogeneity at the type-level in the two samples.

3.1 Descriptive analysis

For our descriptive analysis, we examine each individual's modal choice across the negatively sloped and across the positively sloped budget lines. Negatively sloped budget lines inform us about the amount of money individuals are willing to sacrifice in order to *increase* the payoff of the other individual. In contrast, positively sloped budget lines inform us about the amount of money individuals are willing to sacrifice in order to *decrease* the payoff of the other individual. We focus on the modal choice because it is less susceptible to random responses or to outliers. For each budget line, we label the own-payoff maximizing allocation by $z = 6$, the own-payoff-minimizing allocation by $z = 0$, and the payoff-equalizing allocation by $z = 3$. The other four available allocations on each budget line are equidistantly placed between $0 - 3$ and $3 - 6$, respectively.

We depict the distribution of these modal choices in Figure 1, where an individual's modal choice on positively sloped budget lines is depicted on the x-axis, and her modal choice on the negatively sloped budget lines is depicted on the y-axis. The figure documents two striking patterns: First, while most of the individuals in the general population sample appear to be located in one of three behavioral agglomerations (groups of individuals characterized by a similar behavior), only two behavioral agglomerations appear in the student sample. Second, in the general population sample many decision makers appear to be willing to reduce both their and the other participant's payoff (modal choice on negatively sloped budget lines is lower than $z = 6$), whereas almost no student is willing to do so.

Indeed, the first behavioral agglomeration in the general population sample—located at point (3,3)—comprises subjects who predominantly equalize payoffs both on negatively-sloped budget lines and on positively sloped budget lines. These individuals are characterized by a willingness to decrease both their and the the other's payoff for the sake of equality. The second behavioral agglomeration, located at point (6,3), comprises subjects who predominantly behave in an altruistic way on negatively sloped budget lines but tend to maximize their own payoff on positively sloped budget lines. The third behavioral agglomeration, located at point (6,6), comprises individuals who predominantly maximize their own payoff.

The distribution of modal choices is very different for the student sample, where all but one individual populate the rightmost part of the graph. In this sample, individuals are mainly located in one of *two* behavioral agglomeration. The largest behavioral agglomeration is located at (6,6) and comprises individuals who primarily make self-interested choices.

Then, we also see evidence for another, smaller, behavioral agglomeration at (6,3), which again comprises individuals who predominantly behave in an altruistic way on negatively sloped budget lines but tend to be self-interested on positively sloped budget lines. The remaining individuals are scattered on the rightmost end of the graph, suggesting that there is some individual-level variation in the degree of altruistic concerns for those worse off. Quite strikingly, the largest behavioral agglomeration from the general population, which is located at point (3,3), is absent in the student sample. It therefore appears that virtually no student is willing to decrease both their own and the other's payoff for the sake of equality. This result is very striking. Indeed, despite the fact that our student sample is smaller than our general population sample, *we nevertheless could have expected to find at least 20 to 30 students at point (3,3) if the distribution of students choices corresponded to those of the general population.* In contrast, we find that only one individual populates this point. It is thus extremely unlikely that a much larger sample size would fundamentally change the distribution of preferences in our student sample.

Overall, these two figures provide suggestive evidence that there are stark differences in the distribution of other-regarding preferences types in these two populations.

3.2 Cluster analysis

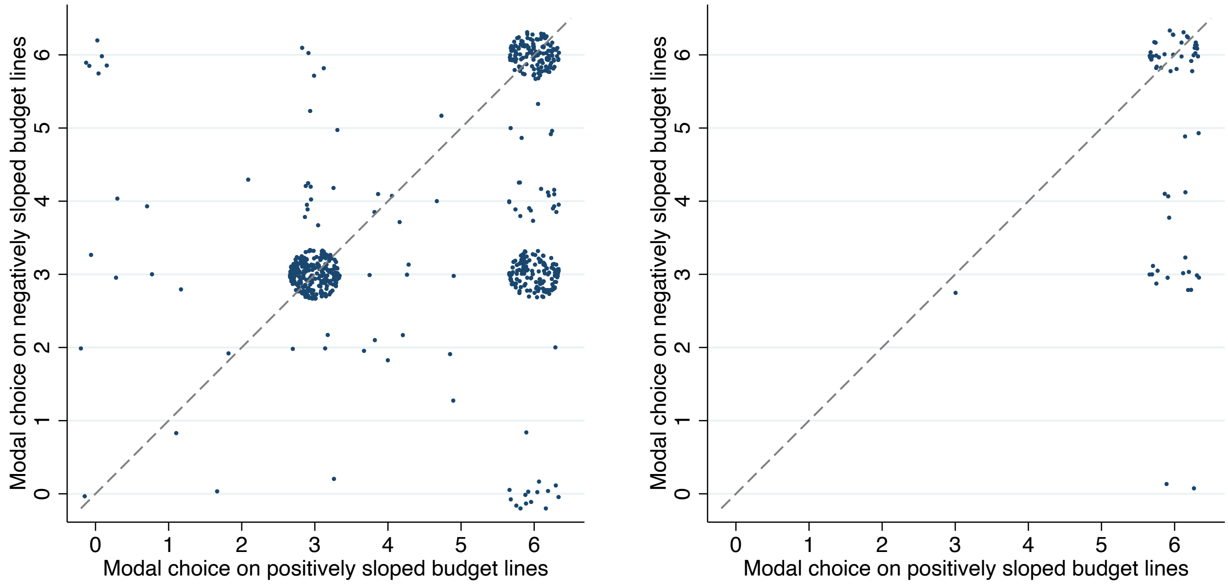
To identify behavioral heterogeneity more systematically, we apply a nonparametric Bayesian approach—the Dirichlet Process (DP) means clustering algorithm (Kulis and Jordan, 2012)—on both samples. This algorithm groups individuals into clusters according to their *behavioral similarities*. In our context, clusters are based on subjects' 12 distributional choices in the money allocation task, and similarity is measured by "how close" an individual's allocation profile is to the average allocation of a cluster.[7] Ultimately, individuals' are assigned to the cluster whose centroid—i.e. the mean allocation in the 12 distributional choices—is the closest to their own allocation profile in the 12-dimensional space of interest. We describe the formalism of the DP-means algorithm in Appendix C.

An important aspect of the DP-means approach is that it enables the identification of preference types without committing to a pre-specified number of different preference types. Moreover, this approach does neither require an ex-ante specification or parameterization of types, nor does it presume a specific error structure. This means that it remains ex-ante

[7]As opposed to the descriptive analysis which used subjects' modal decisions, the DP-means algorithm relies on all 12 choices for clustering, i.e. the algorithm operates in a 12-dimensional space.

Figure 1: Modal choices on negatively sloped and on positively sloped budget lines

(a) General population (b) Students



Note: The figure shows subjects' modal choices among negatively sloped budget lines and among positively sloped budget lines. Each dot represents one individual with unique modes (individuals with multiple modes are excluded from this figure). Dots are jittered in order to make identical modal choices of individuals visible. For each budget line, $z = 6$ indicates an own-payoff maximizing choice, $z = 0$ indicates an own-payoff minimizing choice, and $z = 3$ indicates a payoff-equalizing choice. Note that if we replace individuals' modal choices by their median choices very similar behavioral agglomerations emerge.

agnostic about key distributional assumptions, and it does not constrain heterogeneity to lie within a predetermined set of models or parameter space.[8] The DP-means algorithm allows for all possible type partitions of the data spanning from a representative agent (i.e. a single data-generating process) up to as many types as there are individuals in the population (i.e. n data-generating processes), i.e., it determines the number of preferences types endogenously. Thus, (i) the actual number of types, (ii) the assignment of each individual to one of the types and (iii) the behavioral (preference) properties of the types emerge endogenously.[9]

We run the DP-means algorithm separately on each sample. We display the distribution

[8]In this regard, our approach differs from previous work (e.g. Bellemare et al., 2008; Fisman et al., 2015, 2017; Bruhin et al., 2018) that characterized preference heterogeneity on the basis of structural assumptions on preferences and error terms.

[9]The fact that the number of types adapts to the data has important benefits (see Kulis and Jordan, 2012). Most notably, as previous work has shown (see Comiter et al., 2016), this feature of the algorithm yields higher quality type-separation than methods that specify the number of types prior to clustering (such as k-means).

of types identified in the two samples in the Table 1 below.[10]

Table 1: Type distributions identified using clustering analysis

	General population	**Students**
Inequality averse	50.8%	–
Altruistic	34.4%	40.91%
Selfish	14.8%	54.55%
Other	–	4.56%

In the general population sample, the DP-means algorithm identifies *three* empirically relevant and fundamentally distinct clusters with a clear behavioral interpretation. The largest cluster comprises the 50.8% of the individuals from the general population sample who predominantly equalize payoffs and can therefore be labelled *inequality averse*. The second largest cluster, which accounts for 34.4% of the individuals from the general population sample, comprises individuals who are willing to give up some of their own payoff in order to increase the payoff of those worse off, but are *not* willing to pay to decrease the payoff of those better off, consistent with *altruistic concerns for the worse off*. The smallest cluster comprises the 14.8% of the sample that behave in a *predominantly selfish* way, irrespective of the payoff implications of their choices for the anonymous other participant.

These findings contrast sharply with those of the student sample, where the DP-means algorithm identifies only *two* empirically relevant clusters with a clear behavioral interpretation: one large cluster of subjects who make predominantly self-interested decisions (54.55% of the sample) and another relatively large cluster of subjects behaving in a way consistent with altruistic concerns for the worse off (40.91% of the student sample). Finally, note that there are also three students who are neither assigned to the altruistic nor to the selfish type. For transparency, we bundle these three students together and report them in the "other" category.

4 The Role of Age, Education and Income

In the previous section, we have shown that the distribution of other-regarding types widely differs between the student and the general population sample. A particularly striking re-

[10]Importantly, the DP-means algorithm does *not* assign labels to clusters. In order to do that, one needs to carefully examine subjects' decisions in each cluster. In the Appendix C.2, we show that clusters can be assigned a label with a clear behavioral interpretation.

sult is that the inequality averse type, which comprises about half of the general population sample, entirely disappears in the student sample.

One question that immediately arises is what can explain such a difference, i.e. why are there no inequality averse students?[11] In the following, we explore the role of three factors that could affect subjects' choices in the money allocation task and that systematically vary between the student and the general population samples: age, education, and income.

Because these factors could impact the decisions participants make in the money allocation task, they could explain why the student sample is much more selfish than the general population sample. For example, university students and well educated individuals might make more rational, payoff-maximizing choices as a consequence of their training. Similarly, students might make more own payoff-maximizing choices because of their lower incomes.

In our student sample, all the subjects are young, well educated, and presumably have a relatively low income (if any). In contrast, our general population sample which is representative of the Swiss population is by definition much more diverse: some respondents only finished obligatory school while others have a University degree, some respondents barely make a living while others have extremely high incomes, and the age distribution ranges from 18 to almost 80 years old.

To explore the role of these factors for other-regardingness, we structurally estimate the parameters of an inequality aversion model (Fehr and Schmidt, 1999)

$$V_i\left(w_{ij}\right) = w_{ij}^s - \alpha_i \max\left\{w_{ij}^o - w_{ij}^s, 0\right\} - \beta_i \max\left\{w_{ij}^s - w_{ij}^o, 0\right\}$$

where $w_{ij} = \left(w_{ij}^s, w_{ij}^o\right)$ corresponds to individual i's decision on budget line j on how to allocate money between herself (superscript s for self) and the other person (superscript o for other), α_i denotes aversion towards disadvantageous inequality (behindness aversion) and β_i denotes aversion towards advantageous inequality (aheadness aversion).[12,13] Different pref-

[11] Another question that one could ask is whether the behavior of our student sample is consistent with the behavior of the students in our general poplation sample. Unfortunately, our general population survey did not comprise a question that allows us to unambiguously identify students, but only included a question on labor-market participation (i.e. whether an individual is working full-time, part-time, or is currently out of the labor force). Combining this measure with other variables (such as age and education) would still not allow us to distinguish between University students and other young adults currently out of the labour market (e.g. a young adult who did an apprenticeship and is currently in the military, in a sabbatical, or is unemployed).

[12] We describe the details of the procedure for the structural estimation in Appendix D.

[13] Note that the inequality aversion models of Fehr and Schmidt (1999) is equivalent to Charness and Rabin (2002) in the two person case and in absence of restrictions on the α and β parameters.

erence types are expected to be related to different values of α and β. Inequality averse individuals are both averse to disadvantageous and to advantageous inequality and therefore have $\alpha > 0$ and $\beta > 0$. In contrast, altruistics individuals, whose utility increases in the payoff of the other participant, are characterized by $\alpha \leq 0$ and $\beta \geq 0$ (with at least one of the two parameters being strictly different from zero). Finally, selfish individuals have an α and a β that are both very close to zero since they do not put any weight on the other's payoff.

We document how variations in age, income and education relate to these structural estimates in our general-population sample in Figure 2. Figure 2a shows how variations in age (x-axis) predicts changes in behindness aversion (α, on the left-hand side graph) and aheadness aversion (β, on the right-hand side graph). For clarity, we discretize age into a categorical variable with six categories, where category 1 represents the youngest individuals in our general population sample (aged 18 to 25), and category 6 represents the oldest individuals in our general population sample (aged 65 and more).

These two panels clearly indicate that older individuals generally display much larger degrees of other-regardingness than younger individuals, i.e. they are both more aheadness averse and more behindness averse. For example, individuals that are at least 65 years old have an average estimated α that is about 200 percent larger than the youngest individuals in our sample (two sample t-test: $p < 0.01$), and an average estimated β that is about 40 percent larger than individuals in the youngest age category (two sample t-test: $p < 0.01$). Thus, age appears to be an important driver of differences in other-regardingness between the two samples.

In Figures 2b and 2c , we perform the same exercise for variations in income and in education. While there is no clear relationship between income and other-regardingness (Figures 2b), Figure 2c suggests that education is slightly negatively correlated with other-regardingness, i.e. more educated individuals appear to be less behindness averse (lower α) and less aheadness averse (lower β). However, the magnitude of these effects appears to be much smaller than the age effects.

In Table 2, we further explore the effects of age, income and education on α and β by simultaneously including them in a regression. Consistent with the evidence presented in Figure 2, the main significant predictor of other-regardingness in our general population sample is age, which is strongly associated with both behindness aversion (column 1, $p < 0.01$) and aheadness aversion (column 2, $p < 0.01$). The magnitude of the age effects is large: a 50 years increase in age (e.g. moving from a 20 to a 70 years old) is associated with a increase in

α of about 0.75, and an increase in β of about 0.7. In contrast, higher education is associated with a significant decrease in behindness aversion. For example, respondents in our general population sample who completed university have an average estimated α that is 0.451 points lower than those who only completed obligatory school (column 1, $p < 0.05$). Interestingly, education is not robustly associated with variations in aheadness aversion (column 2). Finally, all the income dummies are statistically insignificant, suggesting that income is not a robust predictor of other-regardingness in our sample.

Another interesting insight from this table is the role of gender: on average, male respondents are characterized by significantly smaller coefficients of other-regardingness than female respondents ($p < 0.01$ for both α and β), suggesting that females are generally more other-regarding in our general population sample. Importantly, these gender effects can *not* explain the observed differences between our student and our general population sample.[14]

Overall, the evidence presented so far indicates that there are stark differences in the distribution of other-regarding preferences between students and the general population, and that these differences can mainly be traced back to age, with a moderate role of education.

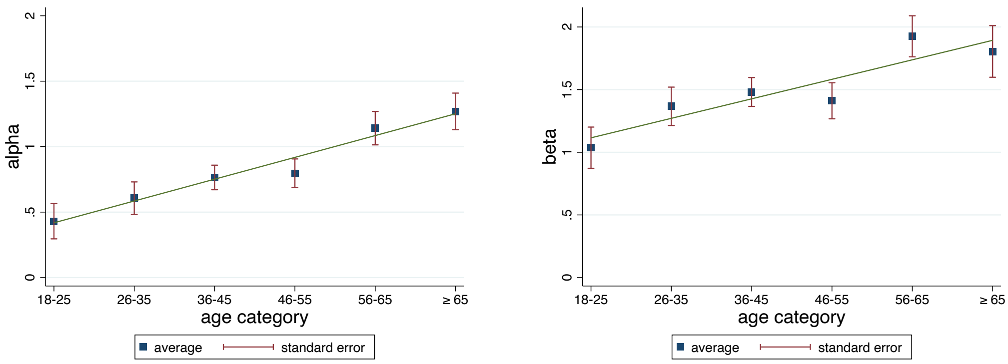
One question that remains open is whether, holding age constant, variations in education can explain part of the variation in α and β. To answer this question, we pool our student sample with the youngest individuals from the general population sample , i.e. respondents from the general population that are aged between 18 and 25—the age of our students. We then regress these subjects' structural parameters on a dummy for being in our student sample and two further controls for age and gender. We report the results in Table 3. This table shows that participants in our student sample are significantly *less* behindness averse (they have a significantly lower α, columns 1-2, $p < 0.01$) and *less* aheadness averse (they have a significantly lower β, columns 3-4, $p < 0.01$) than the non-students—i.e. within young individuals, a higher education mitigates social preferences. In fact, being a student almost completely nullifies α (column 2), and considerably reduces β (column 4).

Altogether, these results suggest that two channels contribute to explaining the absence of an inequality averse type in the students sample. First, younger individuals tend to be generally less other-regarding than older individuals, as characterized by their lower values of α and β. Second, higher education among young individuals (i.e., being a student) is

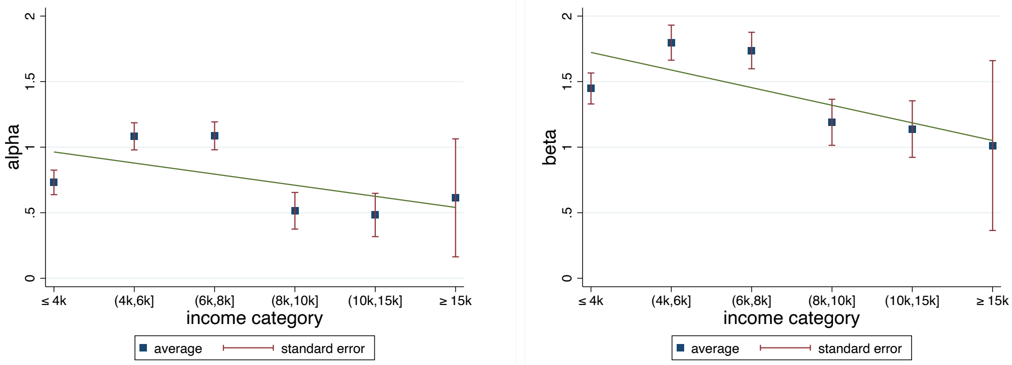
[14]Indeed, there are slightly fewer males in the student sample (49% of male) than in the general population sample (55%). Because males tend to behave more selfishly, we should observe—if anything—*even more selfishness* in the student sample if the proportion of males in the student sample was equal to the proportion of males in the general population sample.

Figure 2: Relating structural estimates of other-regardingess with socio-demographics

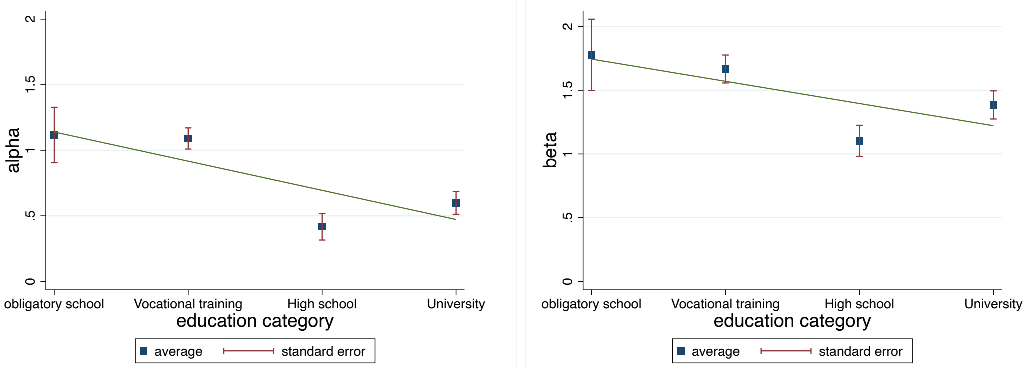
(a) Effect of age on behindness aversion (α) and aheadness aversion (β)



(b) Effect of income on behindness aversion (α) and aheadness aversion (β)



(c) Effect of education on behindness aversion (α) and aheadness aversion (β)



Notes: Dependent variables (y-axis) are the parameter (α, β) estimated using a Hierarchical Bayesian Model with three types.

Table 2: The relationship between other-regardingness and socio-demographics in the general population

	α	β
	(1)	(2)
Age	0.015***	0.014***
	(0.003)	(0.005)
Income bracket: ≤ CHF 4000	-0.167	-0.093
	(0.181)	(0.239)
Income bracket: CHF 4001-6000	0.155	0.262
	(0.187)	(0.249)
Income bracket: CHF 6001-8000	0.211	0.259
	(0.191)	(0.256)
Income bracket: CHF 8001-10000	-0.238	-0.183
	(0.215)	(0.282)
Income bracket: CHF 10001-15000	-0.176	-0.173
	(0.239)	(0.317)
Income bracket: > CHF 15000	-0.228	-0.494
	(0.420)	(0.627)
Education: Vocational training	-0.133	-0.196
	(0.223)	(0.298)
Education: High school	-0.584**	-0.560*
	(0.230)	(0.303)
Education: University	-0.451**	-0.271
	(0.227)	(0.300)
Education: Other	-0.077	-0.047
	(0.251)	(0.331)
Male	-0.593***	-0.718***
	(0.099)	(0.132)
Constant	0.783***	1.459***
	(0.292)	(0.389)
R^2	0.118	0.075
Observations	815	815

Notes: OLS regressions. Dependent variables are the parameter (α, β) estimated using a Hierarchical Bayesian Model with three types. Income brackets and Education brackets are dummie variables indicating respondent's income and highest education, respectively. Omitted categories are individuals who did not disclose their income and those whose highest educational degree is obligatory school, respectively. Age is a continuous variable and male is a dummy varible which equals one if the subject identifies as male. Robust standard errors in parentheses. Levels of significance: $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$

associated with a further large decrease in other-regardingness such that behindness aversion is basically zero among students.

Table 3: The relationship between other-regardingness and socio-demographics among young adults

	α		β	
	(1)	(2)	(3)	(4)
student	-0.629***	-0.597***	-0.773***	-0.746***
	(0.137)	(0.137)	(0.172)	(0.171)
age		-0.005		-0.003
		(0.017)		(0.021)
male		-0.290*		-0.296
		(0.154)		(0.194)
Constant	0.431***	0.665*	1.037***	1.217**
	(0.135)	(0.378)	(0.165)	(0.472)
R^2	0.095	0.116	0.094	0.107
Observations	154	153	154	153

Notes: OLS regressions. For these estimations, we pool the young respondents from the general population sample (aged between 18 and 25) with the student sample. Dependent variables are the parameter (α, β) estimated using a Hierarchical Bayesian Model with three types. Student is a dummy that equals one if the respondent is in our student sample. Age is a continuous variable measuring the respondent's age. Male is a dummy that equals one if the individual identifies as male. Robust standard errors in parentheses. Levels of significance: $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$

5 Discussion

5.1 Are our samples unusual?

Could our results be explained by the fact that our student sample is somewhat unusual relative to other student samples? Or is perhaps our general population sample an outlier? To answer these questions, we collected nine data sets (based on eight different papers) that elicited the distribution of social preferences in student and general population samples. The advantage of these data sets is that they are all based on the Equality-Equivalence Test (Kerschbamer, 2015)—an empirical test that can cleanly disentangle inequality aversion from other forms of social preferences. This test differs from our method in that it requires subjects to make a series of binary choices between an equal-payoff allocation and different alternative allocations in the advantageous and the disadvantageous payoff domain.

Because subjects face the (same) equal-payoff allocation in every single binary choice, the Equality-Equivalence Test (EET) may make this allocation somewhat salient and therefore may favor the emergence of inequality aversion, although this conjecture was not confirmed by the experiments of Krawczyk and Le Lec (2021) who found that the distribution of social preferences is not affected by this feature of the EET. Overall, the EET provides an elegant non-parametric measure of distributional preferences in a non-strategic environment. This has the advantage that confounds arising from strategic considerations are eliminated.[15] Moreover, if the EET is indeed somewhat favoring the emergence of inequality aversion, then it would be an even stronger endorsement of our results if—based on this test—student populations also exhibited rather low levels of inequality aversion.

The distribution of social preferences that emerges from the nine different data sets are presented in Table 4 below. The upper part of the table shows the results from five different student samples with a total of 800 students. The results across the different student samples are remarkably similar.

First, like in our student sample, the share of selfish and altruistic preference types clearly dominate the five student samples of Table 4. As a matter of fact, these two types make up between 80 and 90 percent of the student population. *Second*, between 30 and 58 percent of the students in Table 4 are selfish. Thus, when we compare this to our student sample, the 41% of selfish subjects does not appear to be an outlier. *Third*, in two of the five student samples the altruists comprise roughly 50% of the sample, which is also not very different from our sample where they make up 55%. Thus, while the results from the EET show a somewhat greater share of 7-12% inequality averse students in comparison to our student sample, inequality aversion is far less prevalent compared to the share of selfish or altruistic subjects—just like in our sample.[16]

[15]Note that many early empirical tests of social preferences based on strategic games were ill-equipped to unambiguously disentangle different forms of distributional preferences. For example, rejection rates in the ultimatum game confound inequality aversion with motives such as reciprocity, envy, or spite.

[16]When comparing the clustering in the current paper with the classification in Table 4, it is important to keep in mind that Table 4 is based on the *assumption* that four different preference types (altruistic, inequality averse, envious, selfish) exist, and that each subject can be assigned – based on the choices in the EET – to one of the exogenously assumed types. This differs from the DP-means algorithm used in the current paper which endogenously assigns individuals to behavioral clusters and where the number of clusters is also endogenously determined once the maximally allowable deviation from the centers' clusters is fixed. This explains why—based on the DP-means algorithm—the number of preference types differs from those *assumed* in the EET-classification. Despite these methodological differences, it is remarkable that both approaches lead to the conclusion that (i) altruism and selfishness are far more frequent than inequality aversion among students, and (ii) that inequality

Fourth, Table 4 shows that the share of inequality averse individuals is much larger in the general population than among students. Indeed, the proportion of individuals behaving consistent with inequality aversion ranges from 23 percent in the study by Hedegaard et al. (2021) to about 65 percent in Kerschbamer and Müller (2020) and in the 2018 wave of the German Internet Panel. In addition, these studies also find that the proportion of selfish individuals is systematically lower in the general population than among students.[17] Altogether, these findings suggest that neither our student sample nor our general population sample are unusual.

5.2 Further discussion on the methods

In the previous section, we have shown that the results from our student sample are not unusual if we compare them with other experiments that can cleanly separate inequality aversion from other forms of other-regarding behavior. We now discuss a few other minor differences between our student and our general population samples, and we explain why they are very unlikely to explain our results.

Experimenter demand and subject location. We know that participants in our student sample completed the experiment in our laboratory, but we have little control over where participants from our broad population sample took the study (although we instructed them to complete the study in a quiet place where they would not be disturbed).

Because participants in the student sample were physically present in the laboratory and had to interact with experimenters (e.g. upon their arrival in the lab, or for the payment), one might be concerned that experimenter demand is larger in the laboratory than online.

While we cannot rule out that experimenter demand might be larger in the laboratory, it is important to note that, in the context of our experiment, the "right thing" (or socially desirable) thing to do in our money allocation task would be, if anything, to behave in a *non*-selfish way. Thus, if experimenter demand is larger in the laboratory and if the socially desirable thing to do is to behave in a non-selfish way, then we should expect a lot more selfishness in the general population sample than in the laboratory.

However, we find the opposite: the proportion of subjects behaving in a selfish way is

aversion is vastly less frequent among students compared to the general population.

[17]This finding is also consistent with the studies of Falk et al. (2013), Anderson et al. (2013), and Bellemare et al. (2011, 2008) who report that students are less other-regarding than non-students.

Table 4: Empirical Frequency of Different Distributional Preference Types in nine different Data Sets

Study	Subject Pool	Preference types Altruistic	Inequality Averse	Envy / Spite	Selfish
Student samples					
Kerschbamer (2015)	N = 92 Austria	33.7 %	11.9 %	3.2 %	48.9 %
Krawczyk and Le Lec (2021)	N = 101 Poland	48.5 %	11.9 %	9.8 %	28.7 %
Balafoutas et al. (2012)	N = 132 Austria	28.0 %	8.3 %	5.3 %	58.3 %
Balafoutas et al. (2014)	N = 195 Austria	49.7 %	6.7 %	7.2 %	33.8 %
Paetzel et al. (2014)	N = 280 Germany	29.3 %	9.3 %	1.8 %	58.3 %
Nationally representative samples					
Hedegaard et al. (2021)	N = 885 Denmark	47.1 %	23.2 %	8.6 %	20.0 %
Kerschbamer & Muller (2020) (based on GIP 2016)	N = 2794 Germany	13.4 %	64.8 %	14.0 %	5.0 %
Chapman et al. (2023)	N = 1000 USA	27.5 %	41.9 %	8.3 %	16.6 %
German Internet Panel 2018 (own calculations)	N = 2583 Germany	11.5 %	67.8%	11.0 %	7.9 %

Note: The table is based on papers that elicited distributional preferences with the equality equivalence test (EET), developed by Kerschbamer (2015). However, the reported numbers are the result of our own calculations because the published papers did not allocate individuals to the four different preference types. In terms of the parameters of the Fehr-Schmidt model, the different distributional types are defined as follows: an individual is classified as altruistic if $\alpha \leq 0$ and $\beta \geq 0$ with at least one inequality holding strictly; an individual is inequality averse if $\alpha > 0$ and $\beta > 0$; it is envious if $\alpha \geq 0$ and $\beta < 0$, and selfish if $\alpha = 0$ and $\beta = 0$. GIP 2016 indicates the 2016 wave of the German Internet Panel and GIP 2018 indicates the 2018 wave.

much larger in the student sample. Thus, differential experimenter demand is very unlikely to explain our results. Moreover, note that experimenter demand effects have been shown to be rather small in magnitude, even when explicitly induced (De Quidt et al., 2018).

Opportunity costs of participation. Another possible concern when comparing incentivized behavior in highly diverse samples is that the opportunity costs might differ across subject pools. In our context, one might be worried that the stakes were (perceived to be) larger for students, who tend to have lower incomes than the general population sample.

We believe that this argument cannot be a large issue in our study for the following reasons. First, we slightly adjusted the exchange rate in the money allocation task, from 100 pts = CHF 2.5 in the general population sample to 100 pts = CHF 1.5 in the student sample. This somewhat mitigates the concerns that stakes are (perceived to be) a lot higher in the student sample than in the general population sample. Moreover, participants in both samples were paid approximately what they could have expected given the length of the study (in terms of average pay per hour).[18] It is thus unlikely that opportunity cost of participation in the experiment were (perceived to be) very different across the two samples. Moreover, there is fairly convincing evidence that the stake size does not matter much for behavior in social preference experiments (Fehr et al., 2014; Cameron, 1999; Engel, 2011) unless stakes vary by a factor of 100 or more (Andersen et al., 2011).

Second, our general population sample comprises individuals with highly diverse incomes. This allowed us to explore the extent to which other-regardingness depends on income—a reasonable proxy for opportunity costs. Our results (which we discussed in Section 4) showed that differences in income are *not* associated with differences in social preferences.

Overall, these results are consistent with a recent meta-analysis by Larney et al. (2019) who find an almost zero effect of stake size on ultimatum game offers and a small, although significant, effect of stake size on dictator games offers. Likewise, Engel (2011) also finds a "very small" (but significant) effect of stake size on giving in dictator games in another meta-analysis. The range of stake sizes considered in these meta-analyses is quite large, but the effects are nevertheless very small. It is thus hard to believe that the *large* differences in preferences we documented across our two samples can be explained by a (if any small) difference in opportunity costs.

[18]Both our student and our broad population samples have implicit rules regarding the average hourly pay of subjects. Our incentives were set such that, in both samples, subjects would be approximately compensated accordingly.

Recruitment procedure. The recruitment procedures of our participants also differed between the two samples. While we cannot rule out that recruitment procedure might somewhat affect participation, it is far from clear that this would explain the large differences in preferences documented between our samples. Moreover, we are interested in comparing the *students who are normally taking part in experimental studies* with individuals from the *general population that are normally taking part in studies with broad population samples*. It was thus important, in our view, to address participants from these two samples in the way they are *normally* recruited. For example, it would have been strange to recruit students from our laboratory in a way that differs from the way they are usually recruited. Similarly, it would not have been satisfactory to only survey students from the general population sample, as these are students who might normally *not* participate in laboratory experiments.

Unstable preferences. Another possible concern is related to the fact that the data for the two samples were not collected at the exact same time. For some reason, preferences might have changed between the moment where we collected the general population sample and the moment where we collected data for the student sample.

While we cannot rule out that time has some effect on preferences, it is very unlikely that is can explain such a large change in preferences. Moreover, recent papers that studied the stability of social preferences suggest preferences are rather stable over time, even when these preferences are measured three years apart (Fisman et al., 2023; Fehr ⓡ al., 2023). It is thus very unlikely, in our view, that unstable preferences can explain the large differences we document.

6 Concluding remarks

A large amount of evidence suggests that a substantial share of individuals are other-regarding, i.e., that they not only care about their own payoff but that they also care about the payoff allocated to relevant others. However, the empirical evidence on the existence of other-regarding individuals—or lack thereof—is predominantly based on student samples.

In this paper, we show that the distribution of social preferences in the general population and in the student population fundamentally differ. Using both a descriptive analysis and a more rigourous clustering approach, we document the existence of three qualitatively different behavioral types emerge in the general population (an inequality averse, an altruistic, and a selfish type). In contrast, only two behavioral types emerge in the student sample (the altruistic and the selfish types). The absence of inequality aversion in the student sample is striking, especially when considering the fact that this type comprises more than 50 percent of our general population sample. While the complete absence of inequality aversion in our student sample constitutes a somewhat surprising result, a comparison with other studies that can cleanly separate inequality aversion from other forms of social preferences confirms the conclusion that inequality aversion is vastly less frequent in students compared to the general population.

Using structural estimations, we show that two channels contribute to explaining the absence of an inequality averse type in the students sample. First, younger individuals tend to be generally less other-regarding than older individuals. Second, higher education among young individuals (i.e., being a student) is associated with a further strong decrease in behindness aversion.

Overall, these findings provide a new cautionary tale that, in the domain of social preferences, results from the student population might not extrapolate to the general population because the broader population does not only display more general other-regardingness but also considerably more inequality aversion. These results are important in view of the fact that almost all applications of social preference ideas involve the general population.

References

Abeler, Johannes and Daniele Nosenzo (2015) "Self-selection into laboratory experiments: pro-social motives versus monetary incentives," *Experimental Economics*, 18, 195–214.

Allenby, Greg M and Peter E Rossi (2006) "Hierarchical bayes models," *The handbook of marketing research: Uses, misuses, and future advances*, 418–440.

Andersen, Steffen, Seda Ertaç, Uri Gneezy, Moshe Hoffman, and John A List (2011) "Stakes matter in ultimatum games," *American Economic Review*, 101 (7), 3427–3439.

Anderson, Jon, Stephen V Burks, Jeffrey Carpenter et al. (2013) "Self-selection and variations in the laboratory measurement of other-regarding preferences across subject pools: evidence from one college student and two adult samples," *Experimental Economics*, 16, 170–189.

Balafoutas, Loukas, Rudolf Kerschbamer, Martin Kocher, and Matthias Sutter (2014) "Revealed distributional preferences: Individuals vs. teams," *Journal of economic behavior & organization*, 108, 319–330.

Balafoutas, Loukas, Rudolf Kerschbamer, and Matthias Sutter (2012) "Distributional preferences and competitive behavior," *Journal of economic behavior & organization*, 83 (1), 125–135.

Bellemare, Charles and Sabine Kröger (2007) "On representative social capital," *European Economic Review*, 51 (1), 183–202.

Bellemare, Charles, Sabine Kröger, and Arthur van Soest (2011) "Preferences, intentions, and expectation violations: A large-scale experiment with a representative subject pool," *Journal of Economic Behavior & Organization*, 78 (3), 349–365.

Bellemare, Charles, Sabine Kröger, and Arthur Van Soest (2008) "Measuring inequity aversion in a heterogeneous population using experimental decisions and subjective probabilities," *Econometrica*, 76 (4), 815–839.

Belot, Michele, Raymond Duch, and Luis Miller (2015) "A comprehensive comparison of students and non-students in classic experimental games," *Journal of Economic Behavior & Organization*, 113, 26–33.

Bolton, Gary E and Axel Ockenfels (2000) "ERC: A theory of equity, reciprocity, and competition," *American Economic Review*, 90 (1), 166–193.

Bruhin, Adrian, Ernst Fehr, and Daniel Schunk (2018) "The many faces of human sociality: Uncovering the distribution and stability of social preferences," *Journal of the European Economic Association*.

Camerer, Colin F (2011) *Behavioral game theory: Experiments in strategic interaction*: Princeton university press.

Cameron, Lisa A (1999) "Raising the stakes in the ultimatum game: Experimental evidence from Indonesia," *Economic Inquiry*, 37 (1), 47–59.

Cappelen, Alexander W, Knut Nygaard, Erik Ø Sørensen, and Bertil Tungodden (2015) "Social preferences in the lab: A comparison of students and a representative population," *The Scandinavian Journal of Economics*, 117 (4), 1306–1326.

Carpenter, Jeffrey, Cristina Connolly, and Caitlin Knowles Myers (2008) "Altruistic behavior in a representative dictator experiment," *Experimental Economics*, 11, 282–298.

Carpenter, Jeffrey P, Stephen Burks, and Eric Verhoogen (2005) "Comparing students to workers: The effects of social framing on behavior in distribution games," in *Field experiments in economics*, 10, 261–289: Emerald Group Publishing Limited.

Carpenter, Jeffrey and Erika Seki (2011) "Do social preferences increase productivity? Field experimental evidence from fishermen in Toyama Bay," *Economic Inquiry*, 49 (2), 612–630.

Chapman, Jonathan, Mark Dean, Pietro Ortoleva, Erik Snowberg, and Colin Camerer (2023) "Econographics," *Journal of Political Economy Microeconomics*, 1 (1), 115–161.

Charness, Gary and Matthew Rabin (2002) "Understanding social preferences with simple tests," *Quarterly Journal of Economics*, 817–869.

Comiter, Marcus Zachary, Miriam Cha, HT Kung, and Surat Teerapittayanon (2016) "Lambda means clustering: automatic parameter search and distributed computing implementation."

De Quidt, Jonathan, Johannes Haushofer, and Christopher Roth (2018) "Measuring and bounding experimenter demand," *American Economic Review*, 108 (11), 3266–3302.

Durante, Ruben, Louis Putterman, and Joël Van der Weele (2014) "Preferences for redistribution and perception of fairness: An experimental study," *Journal of the European Economic Association*, 12 (4), 1059–1086.

Engel, Christoph (2011) "Dictator games: A meta study," *Experimental economics*, 14, 583–610.

Falk, Armin, Ernst Fehr, and Christian Zehnder (2006) "Fairness perceptions and reservation wages—the behavioral effects of minimum wage laws," *The Quarterly Journal of Economics*, 121 (4), 1347–1381.

Falk, Armin, Stephan Meier, and Christian Zehnder (2013) "Do lab experiments misrepresent social preferences? The case of self-selected student samples," *Journal of the European Economic Association*, 11 (4), 839–852.

Fehr, Ernst ⓡ Thomas Epper ⓡ Julien Senn (2022) "Other-regarding preferences and redistributive politics," 10.5167/UZH-183052.

——— (2023) "The Fundamental Properties, the Stability, and the Predictive Power of Distributional Preferences."

Fehr, Ernst and John A List (2004) "The hidden costs and returns of incentives—trust and trustworthiness among CEOs," *Journal of the European Economic Association*, 2 (5), 743–771.

Fehr, Ernst and K Schmidt (1999) "A Theory Of Fairness, Competition, and Cooperation," *Quarterly Journal of Economics*, 114 (3), 817–868, `http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=pubmed&cmd=Retrieve&dopt=AbstractPlus&list_uids=9017717075410430622related:nirQ9OZdJX0J`.

Fehr, Ernst, Elena Tougareva, and Urs Fischbacher (2014) "Do high stakes and competition undermine fair behaviour? Evidence from Russia," *Journal of Economic Behavior & Organization*, 108, 354–363.

Fisman, Raymond, Pamela Jakiela, and Shachar Kariv (2017) "Distributional preferences and political behavior," *Journal of Public Economics*, 155, 1–10.

Fisman, Raymond, Pamela Jakiela, Shachar Kariv, and Daniel Markovits (2015) "The distributional preferences of an elite," *Science*, 349 (6254), aab0096.

Fisman, Raymond, Pamela Jakiela, Shachar Kariv, and Silvia Vannutelli (2023) "The distributional preferences of Americans, 2013–2016," *Experimental Economics*, 1–22.

Fisman, Raymond, Shachar Kariv, and Daniel Markovits (2007) "Individual preferences for giving," *American Economic Review*, 97 (5), 1858–1876.

Fréchette, Guillaume R (2016) "Experimental economics across subject populations," *The handbook of experimental economics*, 2, 435–480.

Gächter, Simon, Benedikt Herrmann, and Christian Thöni (2004) "Trust, voluntary cooperation, and socio-economic background: survey and experimental evidence," *Journal of Economic Behavior & Organization*, 55 (4), 505–531.

Gelman, Andrew, John B Carlin, Hal S Stern, and Donald B Rubin (2004) *Bayesian data analysis*, Second Edition: Chapman and Hall/CRC.

Güth, Werner, Carsten Schmidt, and Matthias Sutter (2007) "Bargaining outside the lab–a newspaper experiment of a three-person ultimatum game," *The Economic Journal*, 117 (518), 449–469.

Hedegaard, Morten, Rudolf Kerschbamer, Daniel Müller, and Jean-Robert Tyran (2021) "Distributional preferences explain individual behavior across games and time," *Games and Economic Behavior*, 128, 231–255.

Kerschbamer, Rudolf (2015) "The geometry of distributional preferences and a non-parametric identification approach: The Equality Equivalence Test," *European economic review*, 76, 85–103.

Kerschbamer, Rudolf and D Müller (2020) "Social preferences and political attitudes: An online experiment on a large heterogeneous sample," *Journal of Public Economics*, 182.

Krawczyk, Michal (2011) "What brings your subjects to the lab? A field experiment," *Experimental Economics*, 14, 482–489.

Krawczyk, Michal and Fabrice Le Lec (2021) "How to elicit distributional preferences: A stress-test of the equality equivalence test," *Journal of Economic Behavior & Organization*, 182, 13–28.

Kulis, Brian and Michael I. Jordan (2012) "Revisiting k-means: New Algorithms via Bayesian Nonparametrics," *Proceedings of the 29th International Conference of Machine Learning*.

Larney, Andrea, Amanda Rotella, and Pat Barclay (2019) "Stake size effects in ultimatum game and dictator game offers: A meta-analysis," *Organizational Behavior and Human Decision Processes*, 151, 61–72.

McFadden, Daniel (2001) "Economic choices," *American economic review*, 91 (3), 351–378.

Paetzel, Fabian, Rupert Sausgruber, and Stefan Traub (2014) "Social preferences and voting on reform: An experimental study," *European Economic Review*, 70, 36–55.

Snowberg, Erik and Leeat Yariv (2021) "Testing the waters: Behavior across participant pools," *American Economic Review*, 111 (2), 687–719.

Stoop, Jan, Charles N Noussair, and Daan Van Soest (2012) "From the lab to the field: Cooperation among fishermen," *Journal of Political Economy*, 120 (6), 1027–1056.

Tyran, Jean-Robert and Rupert Sausgruber (2006) "A little fairness may induce a lot of redistribution in democracy," *European Economic Review*, 50 (2), 469–485.

A Background information on the experimental task

A.1 Choice situations in the money allocation task

Figure A.1 depicts the 12 budget lines we use to identify other-regarding preferences in both samples, where the decision maker's own payoff is represented on the x-axis and the recipient's payoff is on the y-axis.[19]

Figure A.1: Budget lines used to identify other-regarding preferences

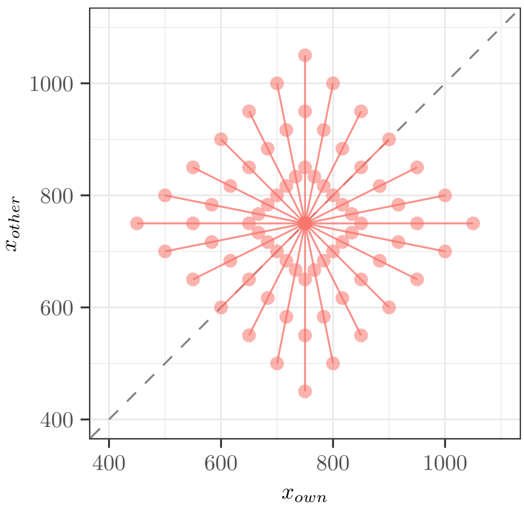


Table A.1 provides further details on these choice situations. The meaning of the list of variables displayed in the Table is as follows:

- 'choiceId': the unique identifier for each choice situation.
- $(own1, other1)$: represents the payoff combination at the lower end of the budget line (in points).
- $(own2, other2)$: represents the payoff combination at the upper end of the budget line (in points).
- 'bundle': indicates to which bundle the respective choice situation belongs to.
- 'slope': the slope of the budget line in the "own payoff – other payoff" space.

[19]These decisions are based on the design in Fehr ⓡ al. (2022).

Table A.1: Choice situations in the money allocation task

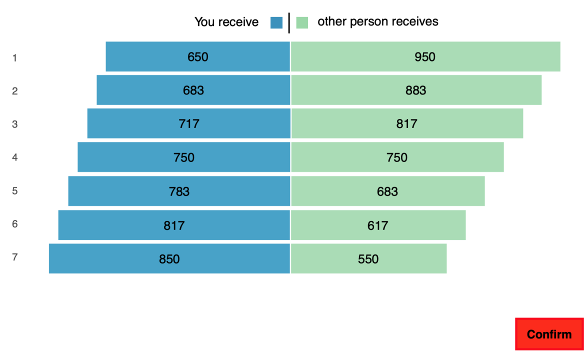
choiceId	own1	own2	other1	other2	slope
1	450	1050	750	750	0.0
2	500	1000	800	700	-0.2
3	550	950	850	650	-0.5
4	600	900	900	600	-1.0
5	650	850	950	550	-2.0
6	700	800	1000	500	-5.0
7	750	750	1050	450	-Inf
8	700	800	500	1000	5.0
9	650	850	550	950	2.0
10	600	900	600	900	1.0
11	550	950	650	850	0.5
12	500	1000	700	800	0.2

A.2 Example choice situation

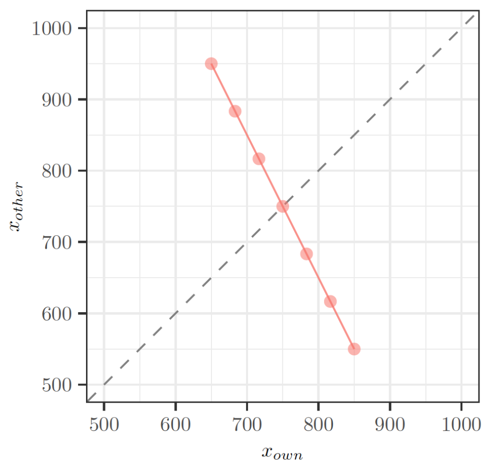
Figure A.2a illustrates how a typical choice situation was presented to our participants. We represented the available choices numerically and graphically in order to make the trade-offs and the associated payoff implications salient. There were always seven interpersonal allocations (labeled by 1 to 7) available per choice situation, and all of them were located on a budget line. Each available allocation consisted of a specific distribution of ECUs between the participant (bars labeled by "You receive") and the other person (bars labeled by "other person receives"). Figure A.2b plots the budget line corresponding to the example depicted in Figure A.2a in the (x_{own}, x_{other})-space ("own payoff", "other's payoff"). In this example, the slope of the budget line is -2, indicating that for every ECU the decision maker gives up, the other player receives 2 ECUs. Perfect equality in payoffs can be achieved by choosing allocation 4.

Figure A.2: Example choice situation

(a) Decision screen



(b) Budget line



B Demographic characteristics of sample populations

In Table B.1, we compare the main sociodemographics of our general population sample to those of the swiss population. In Table B.2, we display the main sociodemographics (age and gender) of our student sample.

Table B.1: Comparison of sample population with the population of Swiss voters

	Sample	Population
Age	46.48	51.08
Male	0.55	0.48
Education : Obligatory school	0.04	0.11
Education : Vocational training	0.37	0.42
Education : High school	0.13	0.10
Education : University	0.34	0.35
Education : Other	0.10	-
Income bracket : ≤ CHF 4000	0.25	0.28
Income bracket : CHF 4001-6000	0.20	0.26
Income bracket : CHF 6001-8000	0.20	0.22
Income bracket : CHF 8001-10000	0.14	0.12
Income bracket : CHF 10001-15000	0.09	0.09
Income bracket : ≥ CHF 15000	0.02	0.03
Income bracket : NA	0.10	-
Unemployed	0.03	0.03
Number of observations	815	

Notes: The table displays descriptive statistics (mean) for the main sociodemographics of the main sample and for the Swiss population. The population data were obtained from the Swiss Federal Bureau of Statistics (2018) and are restricted to the adult Swiss population (i.e. individuals holding a swiss passport who are at least 18 years old).

Table B.2: Student sample

	Mean	S.D.
Age	23.80	4.14
Male	0.49	0.50
Number of observations	66	

Notes: The table displays descriptive statistics (mean and standard deviation) for the student sample.

C Identifying preference types using Dirichlet Process Means

C.1 The method

This appendix provides an overview of the clustering algorithm used to identify the preference types and their distribution in the population. For a more detailed description of the DP-means algorithm and for a discussion of its key differences with other clustering methods such a k-means, see Fehr ⓡ al. (2022, 2023).

Our implementation of the algorithm is based on an iterative refinement. We first span an m-dimensional space, with m denoting the number of budget lines used for the clustering algorithm (in our case, $m = 12$, the twelve budget lines presented in Table A.1 in the main paper). Consequently, each individual's choices are represented by a single point in the 12-dimensional space. We then ask how subjects populate this space. Specifically, we are interested in the number of clusters (i.e. types) that emerge and individuals' assignment to clusters. A cluster is characterized by the set of the individuals assigned to the cluster and the associated mean vector of observations (the "centroid"), which – in our case – represents the mean (cluster- representative) behavior of all individuals in m-dimensional space that belong to the cluster.

We initialize the algorithm with a single centroid specified as the global mean vector. At this stage, all data points are assigned to this single centroid. We then refine by iterating over the following two steps: First, we sequentially go through the list of data points in m-dimensional space (i.e. subjects), and check for each subject whether any of the squared Euclidean distances to the centroid exceeds the cluster penalty parameter λ. If this is the case, we open up a new cluster with the actual data point's location vector as the centroid. Otherwise, we assign the data point to its nearest cluster. Second, we collect the subjects assigned to the same clusters and update the centroids by computing the mean vector for each cluster. These two steps are repeated until convergence is reached, i.e. until there is no change in subjects' assignments.

As Kulis and Jordan (2012) demonstrate, this iterative procedure is equivalent to minimizing the objective

$$\min_{\{g_c\}_{c=1}^{k}} \sum_{c=1}^{k} \sum_{\boldsymbol{x} \in g_c} \|\boldsymbol{x} - \boldsymbol{\mu}_c\|^2 + \lambda k \,,$$

where $\boldsymbol{x}$ denotes the vector of observations, $\boldsymbol{\mu}$ the vector of centroids, and g the cluster partitioning of $\boldsymbol{x}$. It is straightforward to see that this objective is equivalent to the k-means objective except for the additional penalty term λk.

An important aspect of the DP-means approach is that it enables the identification of preference types without committing to a pre-specified number of different preference types. Moreover, this approach does neither require an ex-ante specification or parameterization of types, nor does it presume a specific error structure. This means that it remains ex-ante agnostic about key distributional assumptions, and it does not constrain heterogeneity to lie within a predetermined set of models or parameter space.[20] The DP-means algorithm allows for all possible type partitions of the data spanning from a representative agent (i.e. a single data-generating process) up to as many types as there are individuals in the population (i.e. n data-generating processes), i.e., it determines the number of preferences types endogenously. Thus, (i) the actual number of types, (ii) the assignment of each individual to one of the types and (iii) the behavioral (preference) properties of the types emerge endogenously.[21]

C.2 Behavioral interpretation of the clusters identified in our datasets

C.2.1 General population sample

The application of the DP-means algorithm to the money allocation task in our general population sample suggests the existence of *three* behavioral types. Roughly half of the subjects (50.8%) are assigned to Type 1, around one-third (34.36%) to Type 2, and the remainder (14.85%) to Type 3. The three types differ substantially in terms of their behavior. A careful examination of the decisions of these types permits us to assign them a label with a clear behavioral interpretation.

Figure C.1 shows the distribution of individuals' modal choices among negatively and positively sloped budget lines, separately for each type. The figure thus also enables a judgment regarding how individuals that are assigned to a particular type differ from each other

[20]In this regard, our approach differs from previous work (e.g. Bellemare et al., 2008; Fisman et al., 2015, 2017; Bruhin et al., 2018) that characterized preference heterogeneity on the basis of structural assumptions on preferences and error terms.

[21]The fact that the number of types adapts to the data has important benefits (see Kulis and Jordan, 2012). Most notably, as previous work has shown (see Comiter et al., 2016), this feature of the algorithm yields higher quality type-separation than methods that specify the number of types prior to clustering (such as k-means).

and how large these deviations are. The figure shows that the vast majority of individuals in type 1 make modal choices that are payoff-equalizing (z = 3), and they do so for both the negatively sloped budget lines (Figure C.1a) and the positively sloped budget lines (Figure C.1b). They thus exhibit a willingness to pay (i) for reducing inequality when this involves increasing the other individual's payoff (i.e., for negative slopes) and (ii) when it involves decreasing the other individual's payoff (i.e., for positive slopes). For this reason, we assign the label *inequality averse* to type 1—which comprises 50.8% of our sample.

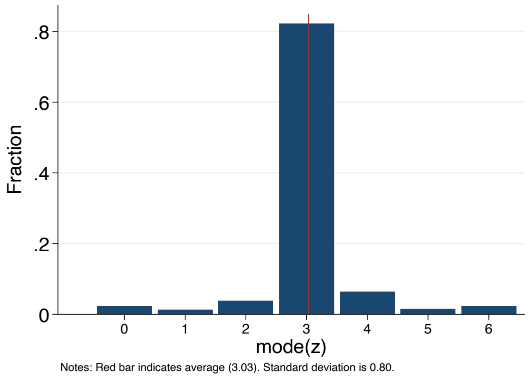
This pattern contrasts sharply with the individuals assigned to type 2 and type 3. Individuals assigned to type 3 (see Figure C.1e and C.1f), in particular, deviate sharply from the inequality averse type: in the vast majority of the cases their modal choice is the own-payoff maximizing (z = 6) allocation regardless of whether budget lines have a positive or a negative slope. These 14.8% of individuals can therefore be characterized as *predominantly selfish*. Finally, individuals assigned to the type 2 cluster differ sharply from the inequality averse type for positively sloped budget lines where the own-payoff (and simultaneously other-payoff) maximizing allocation is basically their modal choice in 100% of the cases (see Figure C.1d). However, the behavior of type 2 individuals for the negatively sloped budget lines resembles that of the inequality averse individuals because the egalitarian allocation is their modal choice in roughly 60% of the cases (Figure C.1c). Thus, these individuals are *willing to increase* other individuals' payoff in the domain of advantageous inequality, i.e., when they are better off than others, but they are *never willing to reduce* other individual's payoff on positively sloped budget lines to avoid disadvantageous inequality. We therefore label individuals belonging to this type, 34.4% of our population, as subjects with an *altruistic concern*. The label "altruistic" is due to their willingness to sacrifice money to mitigate advantageous inequality and help those worse off.

Another remarkable aspect of Figure C.1 is that there is generally very little within-type variation, as indicated by the low standard deviation associated with each of the graphs shown in the figure. This low within-type variation provides a further justification for speaking of different types of preferences; and the fact that the typical choices of the three types sharply differ justifies the notion that the preference differences across types are of a fundamental nature.

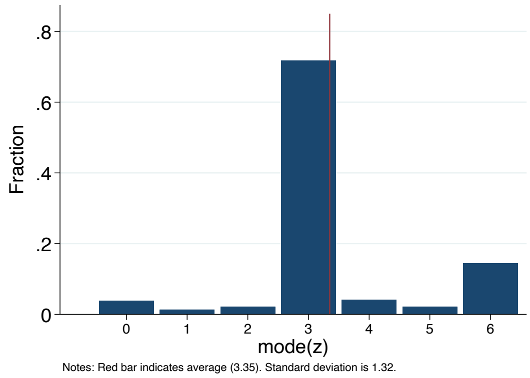
If our preference interpretation of the behavioral types is correct and stable across budget bundles, the different types should display characteristic behavioral patterns in decision situations that lie above or below the 45-degree line. In other words, the inequality averse

Figure C.1: Distribution of individuals' modal choices for each preference type (General Population sample)

Type 1 : Inequality averse (50.8%)

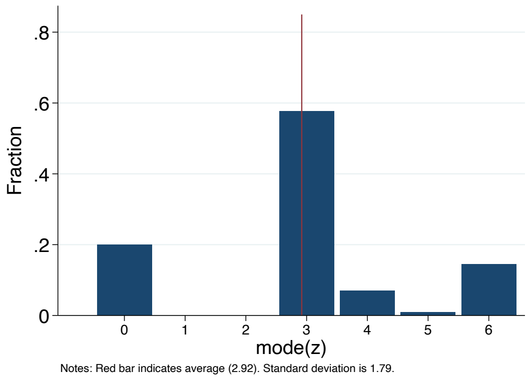


(a) Negatively sloped budget lines

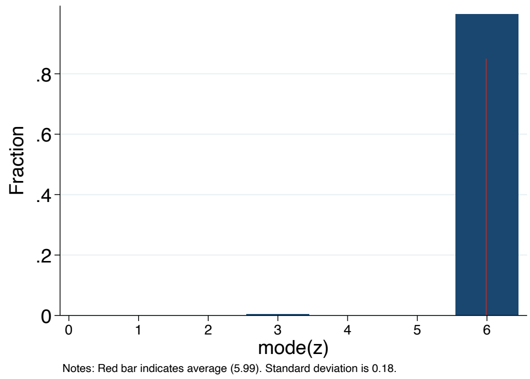


(b) Positively sloped budget lines

Type 2 : Altruistic (34.4%)

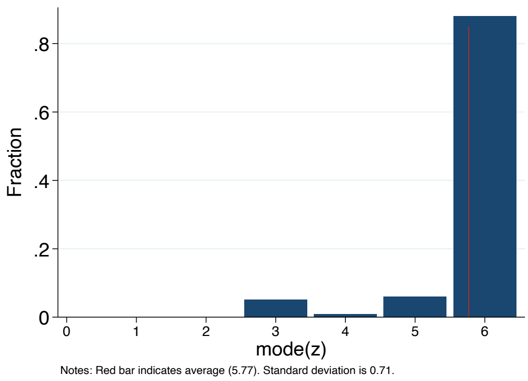


(c) Negatively sloped budget lines

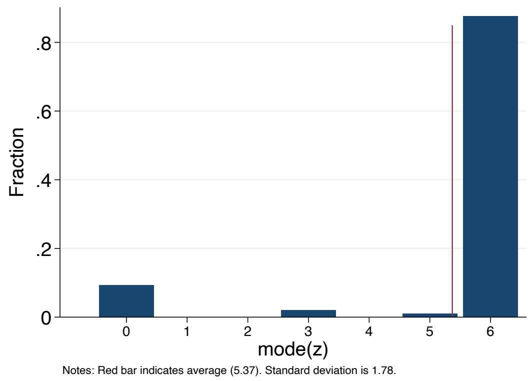


(d) Positively sloped budget lines

Type 3 : Predominantly selfish (14.8%)



(e) Negatively sloped budget lines



(f) Positively sloped budget lines

Note: The figure shows the distribution of individuals' modal choices among negatively sloped and among positively sloped budget lines for each of the three behavioral types identified by the clustering algorithm. For each budget line, $z = 6$ indicates an own-payoff maximizing choice, $z = 0$ indicates an own-payoff minimizing choice, $z = 3$ indicates a payoff-equalizing choice. The red vertical line indicates always the average over all modal choices.

type should also display a preference for equality in these decision situations. Likewise, the selfish type should also predominantly maximize its own payoff in these alternative budget lines. In the Appendix material of Fehr ⓡ al. (2022), we show that this is indeed the case. We also show that taking the median instead of the mode leads to an identical behavioral characterization.

C.2.2 Student sample

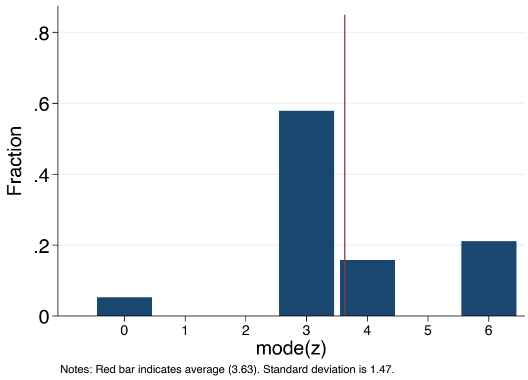
The application of the DP-means algorithm to the money allocation task in our student sample suggests the existence of *two* behavioral types. 40.91% of the students are assigned to Type 1, while 54.55% are assigned to Type 2. The remaining three subjects are each assigned to a separate behavioral cluster. Because these clusters are populated by only one individual each, they cannot be considered empirically relevant behavioral types. We therefore do not display their behavior in this Appendix.

Figure C.2 shows that the vast majority of students assigned to type 1 make modal choices that are consistent with the altruistic type uncovered in the general population sample (Type 2, Figure C.1c and d): they are *willing to increase* other individuals' payoff in the domain of advantageous inequality, i.e., when they are better off than others, but they are *not willing to reduce* other individual's payoff on positively sloped budget lines to avoid disadvantageous inequality. We therefore label them as 'altruistic' in light of their willingness to sacrifice money to mitigate advantageous inequality and help those worse off.

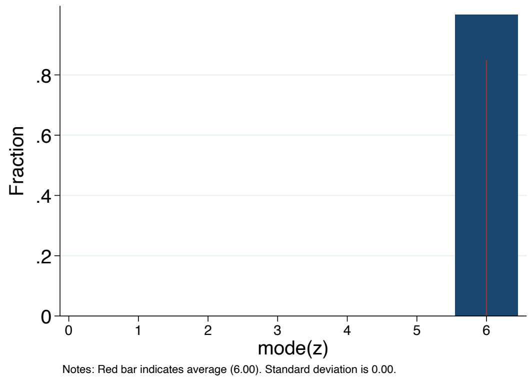
In contrast, students assigned to type 2 (panel c and d) make predominantly own-payoff maximizing decisions: in the vast majority of the cases their modal choice is the own-payoff maximizing allocation regardless of whether budget lines have a positive or a negative slope. We therefore label these individuals as being *predominantly selfish*, just like the predominently selfish type documented in the general population sample (Type 3, Figure C.1e and f).

Figure C.2: Distribution of individuals' modal choices for each preference type (Student sample)

Type 1: Altruistic (40.91 %)

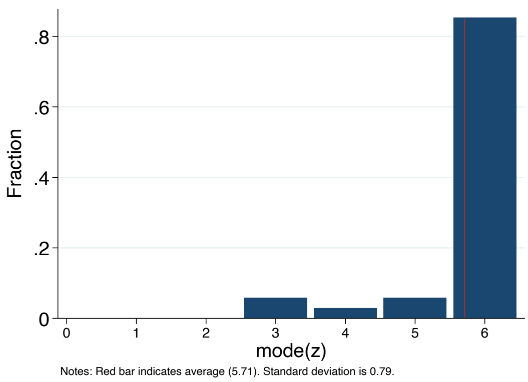


(a) Negatively sloped budget lines

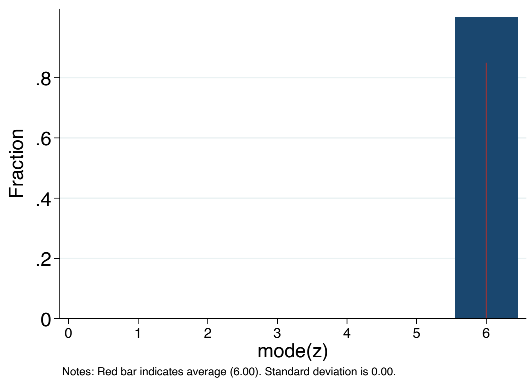


(b) Positively sloped budget lines

Type 2 : Predominantly selfish (54.55 %)



(c) Negatively sloped budget lines



(d) Positively sloped budget lines

Note: The figure shows the distribution of individuals' modal choices among negatively sloped and among positively sloped budget lines for each of the three behavioral types identified by the clustering algorithm. For each budget line, $z = 6$ indicates an own-payoff maximizing choice, $z = 0$ indicates an own-payoff minimizing choice, $z = 3$ indicates a payoff-equalizing choice. The red vertical line indicates always the average over all modal choices.

D Material related to the structural estimation of other-regarding preferences

This appendix provides an overview of the structural estimation approach used to recover individual-levels parameters of behindness aversion (α) and aheadness aversion (β). We provide a more detailed treatise of this approach and discuss the out-of-sample predictive ability of such estimates in Fehr ⓡ al. (2023).

We estimate individual parameters of an inequality aversion model with the 12 choice situations (budget lines) used for type identification. For each budget line j (see Figure A.1 in the main text) presented as a separate choice situation, the individual i chooses one out of seven possible allocations between herself (superscript s for self) and another person (superscript o for other), denoted by $w_{ij} = \left(w_{ij}^s, w_{ij}^o\right)$. As usual, we assume that the individual seeks to maximize her utility.

To estimate a model that permits individuals to differ in their preferences, we need to make explicit assumptions about the specification of (i) the deterministic component, (ii) the stochastic choice model, and (iii) the nature of unobserved heterogeneity in our data.

We estimate the parameters of an inequality aversion model (Fehr and Schmidt, 1999). This (deterministic) model assigns value V_i to each interpersonal distribution w_{ij}, such that

$$V_i\left(w_{ij}^s, w_{ij}^o\right) = w_{ij}^s - \alpha_i \max\left\{w_{ij}^o - w_{ij}^s, 0\right\} - \beta_i \max\left\{w_{ij}^s - w_{ij}^o, 0\right\},$$

where α_i denotes aversion towards disadvantageous inequality (behindness aversion) and β_i denotes aversion towards advantageous inequality (aheadness aversion).

To make the model operational, we assume a random utility model (McFadden, 2001). That is, we also estimate an idiosyncratic error parameter $\zeta_i > 0$ in addition to the two behavioral parameters α_i and β_i, such that the value of an interpersonal allocation w_{ij} depends on three parameters, i.e. $V_i\left(w_{ij}^s, w_{ij}^o\right) := V(\cdot, \alpha_i, \beta_i, \zeta_i)$. The choice model yields for each allocation a choice probability

$$\text{Prob}\Bigg(V_i\left(w_{jk}^s, w_{jk}^o\right) - V_i\left(w_{j'k}^s, w_{jk}^o\right)\Big) > \varepsilon_{j'k} - \varepsilon_{jk}\Bigg) = \frac{e^{\zeta_i V_i\left(w_{jk}^s, w_{jk}^o\right)}}{\sum_l e^{\zeta_i V_i\left(w_{lk}^s, w_{lk}^o\right)}}.$$

For simplicity, let $\theta' = \{\alpha, \beta, \zeta\}$ denote the parameter vector being estimated, and ignore the individual parameter indices. The probability of observing a vector of allocations $z_i = (z_{i1}, z_{i2}, ...)$ for individual i, conditional on θ_i' and ζ_i, is

$$L(w_{ijk}^s, w_{ijk}^o|\theta') = \prod_k \prod_j \left(\frac{e^{\zeta_i V_i\left(w_{jk}^s, w_{jk}^o\right)}}{\sum_l e^{\zeta_i V_i\left(w_{lk}^s, w_{lk}^o\right)}} \right)^{\mathbb{1}\left[z_{ik}^* = z_{jk}\right]}$$

The likelihood *not conditional* on the individual parameters is the integral of $L(z_i|\theta')$ over all θ', i.e.

$$L\left(z_i|\mu, \Omega\right) = \int L(z_i|\theta')\phi\left(\theta'|\mu, \Omega\right) d\theta'$$

where $\phi(\theta'|\mu, \Omega)$ is the normal density with mean μ and variance Ω.

The posterior distribution Q of μ, Ω is

$$Q(\mu, \Omega|z) \propto \prod_i L\left(z_i|\mu, \Omega\right) P\left(\mu, \Omega\right)$$

where $z = (z_{11}, z_{12}, ..., z_{ik}, ..., z_{nm})$, L is the likelihood and P is the prior distribution.

We employ a Bayesian hierarchical approach to estimate the model parameters.[22] To do this, we assume a wide prior distribution, and draw from the posterior using a Gibbs sampler, with the model parameters following a multivariate normal distribution with appropriate parameter constraints.

[22]The procedure is described in detail in Allenby and Rossi (2006) and Gelman et al. (2004). Note that the hierarchical Bayes approach still asssumes a partial pooling, and not no pooling as individual estimates for each subject in the sample. More specifically, it estimates mean and standard deviation of the population parameter distribution. Each individual is part of this distribution and is allowed to deviate from the population mean. However, the model disciplines individuals to not depart "too strongly" from population-typical behavior. It is a feature of these models that individuals who show rather erratic behavior, or behavior departing strongly from typical behaviors, appear to be closer to the population mean (this is also called shrinkage).