Journal of Public Economics 259 (2026) 105681

Contents lists available at ScienceDirect

Journal of Public Economics

journal homepage: www.elsevier.com/locate/pubec

# Choosing (not) to choose: Uncovering intrinsic preferences for choice autonomy☆

Jana Freundt[a], Holger Herz[b,*], Leander Kopp[b]

[a] Lucerne University of Applied Sciences and Arts, School of Social Work, Werftestrasse 1, 6002 Luzern, Switzerland
[b] University of Fribourg, Department of Economics, Boulevard de Pérolles 90, 1700 Fribourg, Switzerland

## HIGHLIGHTS

- We show the existence of intrinsic preference for choice autonomy independent of an instrumental benefit to choice.
- We reveal preference heterogeneity with a majority valuing choice autonomy and a substantial minority being averse to it.
- We introduce a novel elicitation method to measure intrinsic preferences for choice autonomy.
- The elicitation methods elicits and controls for inconsistences in choices.

## ARTICLE INFO

*Keywords:*
Preferences for Choice Autonomy
Preferences for Randomization
Preference Measurement
Dynamically Optimized Sequential Experiment

## ABSTRACT

Personal autonomy has been argued to be fundamental to well-being and is often discussed as an important driver of economic and political behavior. However, preferences for autonomy remain poorly understood. We propose a novel elicitation method for an individual's intrinsic preference for choice autonomy that fully controls for instrumental value benefits of choice. We show that, while a majority of subjects in our experiment display a preference for choice autonomy, a significant minority is averse to their own choice.

## 1. Introduction

Understanding individual preferences is fundamental to properly model and predict economic behavior. As a consequence, a vast economic literature has empirically studied and substantially advanced our understanding of individual preferences, particularly in the domains of risk, time, and social behavior.

In this paper, we study intrinsic preferences for choice autonomy. Arguments for such preferences can for example be found in self-determination theory (Deci and Ryan, 1985), which hypothesizes that autonomy is "essential for ongoing psychological growth, integrity and well-being" (Deci and Ryan, 2000, p.229) and in the capabilities approach by Sen and Nussbaum (Sen, 1985; Nussbaum, 2000), which emphasizes that freedom of choice, and not only outcomes, is important for a person's quality of life. These theoretical approaches further develop early ideas on liberty such as John Stuart Mill's argument for liberty as "one of the elements of wellbeing" (1859, Chapter III).

These arguments contrast with the dominant view in economics, which focuses on the instrumental benefits of choice autonomy, derived

☆ We thank the audiences at the 2022 World Economic Science Association Conference, the 2022 North American Summer Meeting of the Econometric Society, the Behavioral Measurement Conference 2022 (in particular our discussant Michael Kosfeld), the Annual Congress of the Swiss Society of Economics and Statistics 2022, the ECONtribute Workshop on Autonomy and Paternalism 2022, the 12th International Conference of the French Association of Experimental Economics, the Foundations of Utility and Risk Conference 2022, the European ESA Meeting 2022, the Workshop on Paternalism at the Max Planck Institute for Tax Law and Public Finance 2022, the iSee workshop at NYU Abu Dhabi 2023, the behavioral and experimental economics seminar at the University of Zurich, the microeconomics seminar in Konstanz, the IIPF Annual Conference 2023 (in particular our discussant Matthias Rodemeier), the CESifo Behavioral Economics Conference 2023 as well as Arno Apffelstaedt, Berno Buechel, Alexis Ghersengorin, Yoram Halevy, Menusch Khadjavi, Florian Schneider, Christian Thöni, Stephanie Wang, Leeat Yariv and Christian Zihlmann for helpful comments. We thank Kay Waefler for excellent research assistance. The project has received ethics approval from the IRB at the Department of Psychology as well as the IRB at the department of Management, Economics and Social Sciences at the University of Fribourg. We gratefully acknowledge financial support from ERC Starting Grant 2018-803332 CONTROL.

* Corresponding author.
*Email addresses:* jana.freundt@hslu.ch (J. Freundt), holger.herz@unifr.ch (H. Herz), leander.kopp@hotmail.com (L. Kopp).

https://doi.org/10.1016/j.jpubeco.2026.105681
Received 12 May 2025; Received in revised form 20 May 2026; Accepted 22 May 2026
Available online 3 June 2026

from the ability to maximize utility through choice.[1] Potential intrinsic value components of autonomy have been largely ignored, despite them potentially being a crucial determinants of economically relevant behavior. Studies that discuss their potential relevance usually treat them as a residual theory, for example to explain wage differentials in self-employment, entrepreneurship, and science (Hamilton, 2000; Stern, 2004; Hurst and Pugsley, 2011; Astebro et al., 2014) or as an explanation for the underdelegation of decision rights in organizations (Fehr et al., 2013; Fehrler and Janas, 2021). However, assigning residuals to a preference poses an identification challenge, because they could be the result of measurement error in relevant control variables (Gillen et al., 2019) or due to omitted variable bias. The underlying preference ultimately remains unidentified.

The identification of intrinsic value components of choice autonomy using observational data or survey measures is, to the best of our knowledge, infeasible. Whenever an individual faces a choice or is asked about the importance of having choice autonomy, as is done in commonly used surveys that aim to measure perceptions and attitudes towards autonomy, there are likely instrumental benefits associated with the choice or question. Importantly, due to heterogeneity in preferences over outcomes or differences in beliefs, there can be large individual heterogeneity in the (subjectively perceived) instrumental value component of autonomy. Consequently, controlling for the instrumental benefits of choice at the individual level is essential.

Given the complexity and multidimensionality of the concept of personal autonomy, any empirical measure of an underlying preference must necessarily be based on a precise behavioral definition that cannot do justice to all aspects related to personal autonomy. In this paper, we focus on intrinsic preferences for choice autonomy, defined as a *"desire to make decisions oneself rather than having someone else decide over one's own consequences on one's behalf—independent of the instrumental utility associated with the decision."* Other papers have studied preferences for related concepts, such as preferences for decision rights (Bartling et al., 2014; Ferreira et al., 2020; Buffat et al., 2023), where decision rights denote the right to implement effort in a project selection task that had consequences for multiple parties, preferences for having the ability to influence others (Pikulina and Tergiman, 2020b; Gomez-Martinez and Herz, 2025), or preferences for controlling one's own payoffs and non-interference (Owens et al., 2014; Neri and Rommeswinkel, 2016; Sloof and von Siemens, 2017; Pikulina and Tergiman, 2020a; Boissonnet and Ghersengorin, 2022; Meemann, 2023).[2]

The approach proposed in this paper moves beyond the above listed papers by (i) focusing on choice autonomy (as defined above), and by (ii) providing a novel and much simpler methodology in which measurement is directly based on an expressed willingness to pay to take an active choice, instead of an ex-post evaluation of the monetary consequences of different outcomes. Our elicitation tool also provides a significant advancement over existing related measures by providing a non-complex decision environment for subjects that can be applied in a wide variety of settings, including outside of an experimental laboratory, which is much better suited for the study of heterogeneity within as well as the behavioral consequences of the measured preference.

Furthermore, (iii) our approach allows better capturing of individual heterogeneity in preferences for choice autonomy, which remains largely unexplored in the above-mentioned literature. Although our primary motivation suggests that autonomy is positively valued by individuals, individual differences might be substantial. In fact, other evidence suggests that some people may exhibit an *aversion* to making their own choices. Agranov and Ortoleva (2017), Dwenger et al. (2018), and Cettolin and Riedl (2019) show that some individuals display a preference to delegate choices between lotteries to a randomization device, even at a cost, thus opting *not* to choose themselves. Agranov and Ortoleva (2017) argue that such behavior is an expression of preferences for randomization. We discuss the connection between our approach and preferences for randomization below and in Section 6.

Our preference elicitation tool relies on a simple two-step procedure: First, a choice set is identified that contains only alternatives between which an individual is revealed indifferent. Second, an individual's willingness to pay to make a choice from the choice set herself, rather than having someone else choose on her behalf, is elicited. Identification of intrinsic preferences for autonomy follows from the revealed indifference between the choice alternatives elicited in step 1, which implies that there is no instrumental benefit in choosing oneself, as the instrumental utility received from all alternatives is constant, and thus independent of choosing oneself. Therefore, only if an individual has an intrinsic preference for (or aversion to) choice autonomy, she should display a positive (or negative) willingness to pay.

The principle underlying our elicitation tool can be used for any type of alternatives over which individuals have well-defined preferences; it can thus be applied across a large variety of contexts. In this application, we use lotteries as alternatives.[3] In step 1, to identify a subject's point of indifference, subjects make 10 binary decisions between lotteries, where one lottery remains fixed throughout while the high payoff of the other lottery is adjusted from decision to decision. For step 2, a choice set is constructed that contains two lotteries such that the individual is expected to be revealed as indifferent between them, based on the information gained in step 1. We then elicit the willingness to pay to make a choice from this choice set herself, rather than having someone else make this choice on the subject's behalf.

Given that identification critically relies on indifference between alternatives in step 2, it is important to obtain information on how well indifference is identified. A caveat of many methods used to elicit indifference (Becker et al., 1964; Holt and Laury, 2002) is that they only deliver reliable estimates if individuals are *consistent in their choices*. Instead of assuming consistency, we control for potential bias by jointly maximizing the information gain about an individual' point of indifference and her choice consistency. To this end, we designed both step 1 and step 2 as a 'Dynamically Optimized Sequential Experiment' (DOSE, Wang et al., 2010; Chapman et al., 2022, 2024), which delivers an estimate of an individual's indifference point and her willingness to pay, respectively, together with an estimate of an individual's choice consistency.

Based on data from two large-scale online experiments on Prolific.co with a total of 1422 individuals from around the globe, we find that, on average, individual willingness to pay for choice autonomy is significantly larger than zero. On average, the intrinsic value component accounts for 5.0% of the overall expected value generated by the decision. We also find substantial heterogeneity in preferences for choice autonomy. While 53.8% of our subjects have a strictly positive willingness to pay, 19.1% have a willingness to pay of zero, and the willingness to pay is strictly negative for 27.1% of subjects. Those with a positive willingness to pay are on average willing to pay 14% of the expected value of the lottery to retain choice autonomy. Those with a negative willingness to pay are on average willing to pay 9% of the expected value of the lottery to delegate.

To assess the robustness of our findings, we use information on choice consistency at the individual level. First, 49% of our sample shows highly

[1] An exception is Frey et al. (2004), who argued that independence and autonomy at the workplace are sources of procedural utility that raise happiness.

[2] There is also a related literature on (i) the utility consequences of the size of choice sets (Sethi-Iyengar et al., 2004; Iyengar and Kamenica, 2010; Scheibehenne et al., 2010; Fabrice and Tarroux, 2020), (ii) the desirability of the consequences of choice (Botti et al., 2009; Bobadilla-Suarez et al., 2017; Bartling and Fischbacher, 2012), and (iii) on agency and socio-economic background (Snibbe and Markus, 2005).

[3] Lotteries are particularly suitable because they allow for incremental adjustment of value, which is essential to closely approximate an individual's point of indifference. At the same time, we recognize that lotteries are just one instantiation, and intrinsic preferences for choice autonomy may be highly context dependent, thereby limiting the immediate generalizability of our setup.

consistent choice patterns and thus have precisely identified indifference points. When conditioning on this subset, the average willingness to pay for choice autonomy remains similar at 5.6% of the expected value generated by the decision. Second, utilizing structural assumptions on utility functions, we can calculate how large any potentially remaining instrumental benefit of choice autonomy is at the individual level. For highly consistent subjects, this residual is very small, accounting for less than 1% of the overall utility generated by the lotteries. It also remains small for those individuals who display choice inconsistencies in step 1. Thus, imprecise measurement of the indifference set cannot explain the existence of a willingness to pay for choice autonomy in our setting.[4]

Having established the existence of intrinsic preferences for autonomy in the context of choices over lotteries, we assess their relation to well-established related constructs in psychology that measure different aspects of people's perceptions of their internal control and freedom of choice (autonomous functioning Deci and Ryan, 2006, locus of control Rotter, 1966, self-efficacy Schwarzer et al., 1995, desirability of control Burger and Cooper, 1979, and the World Values Survey question on perceived freedom and control Inglehart et al., 2014). We find that our preference measure is unrelated to these constructs, which points to a fundamental conceptual difference between measures of *perceptions of autonomy* and our measure of intrinsic preferences.

To assess robustness, we replicate our experiment in the laboratory with university students. The data confirm the original results; both the average willingness to pay and its distribution are quite similar to the online setting. To shed further light on the potential determinants of preferences for choice autonomy, we conducted a "COIN" treatment, in which participants could delegate to an objective coin toss instead of a human delegate. In this treatment, the average willingness to pay remains significantly positive but drops by half. 46.7% of participants still exhibit a positive willingness to pay. This difference suggests that intrinsic preferences for choice autonomy may be moderated by the identity of the delegate, in this case whether it is a human or a machine (random device).

Our motivation is to understand why individuals intrinsically value retaining choice, based on psychological utility derived from agency and self- determination. This stands in contrast to the literature on preferences for randomization and deliberate stochasticity, which is motivated by the opposite observation: that individuals may not want to choose directly and instead prefer to randomize between lotteries—a behavior that Machina (1985) and Cerreia-Vioglio et al. (2019) rationalize through preferences that are convex over lotteries. A negative willingness to pay in our experiment could similarly be rationalized by randomization preferences, when viewing the delegate merely as a randomization device. Conversely, a preference for choosing oneself corresponds to an aversion to randomization in this literature, implying preferences that are concave over lotteries, as also discussed by Machina (1985) and Cerreia-Vioglio et al. (2019). However, this discussion has remained purely theoretical, as identifying such "randomization-averse" behavior requires choices over alternatives between which individuals are indifferent—an empirical setting that has not existed until now. Our data allow such identification and, when viewed through the lens of randomization preferences, suggest that aversion to randomization is more prevalent than preferences for randomization.

Which of these two mechanisms ultimately drives behavior—preferences for choice autonomy or preferences for randomization—cannot be resolved by our experiment. Importantly, both mechanisms can manifest themselves in revealed preferences for (or against) choice autonomy. Whether one mechanism dominates the other, or whether they coexist within or across individuals, remains an open question for future research.

## 2. Theoretical framework

To motivate our behavioral elicitation method for intrinsic preferences for choice autonomy, we first provide a general choice theoretic framework that guides the elicitation tool.

### 2.1. General setup

We consider an individual who faces a two-stage decision problem over a set of alternatives $x \in C$. In the first stage, she chooses a decision mode $c \in \{0, 1\}$, where $c = 1$ denotes retaining the choice from $C$ herself, and $c = 0$ denotes delegation. Retaining the choice is associated with paying a price $p \in \mathbb{R}$, while delegation is costless ($p = 0$).

The individual has von Neumann–Morgenstern preferences over pairs $(x, p)$ under each decision mode, represented by a utility function $V(x, p, c)$. The decision mode $c$ may affect utility either (i) indirectly through the realization of $x$ and $p$, (ii) directly, capturing intrinsic procedural utility stemming from choice autonomy, or (iii) directly through an interaction of $c$ with $x$, capturing potential interactions between a chosen alternative and the choice process.[5] Further, we assume that utility decreases with the price $p$[6]:

$$\frac{\partial V(x,p,c)}{\partial p} < 0 \quad \text{for all } x \in C,\ c \in \{0,1\}.$$

If the individual chooses $c = 1$, she selects an alternative

$$x^*(1) \in \arg\max_{x \in C} V(x, p, 1),$$

yielding indirect utility

$$W(1) = \max_{x \in C} V(x, p, 1).$$

If she chooses $c = 0$, the alternative is selected by a delegate. Let $\pi(x \mid c = 0)$ denote the individual's subjective belief over which alternative the delegate will choose, giving utility under delegation[7]:

$$W(0) = \sum_{x \in C} \pi(x \mid c = 0)\, V(x, 0, 0).$$

The optimal first-stage choice is then given by

$$c^* \in \arg\max_{c \in \{0,1\}} W(c).$$

To further structure the framework, it is useful to decompose utility into a purely consequentialist component and a residual term capturing any direct effect of the decision mode. Let $\tilde{V}(x, p)$ denote a benchmark utility that depends only on outcomes $(x, p)$. Without loss of generality, we can write

$$V(x, p, c) = \tilde{V}(x, p) + \Delta(x, p, c), \tag{1}$$

where $\Delta(x, p, c)$ captures all non-consequentialist components of utility. For normalization, we set $\Delta(x, p, 0) = 0$ for all $(x, p)$, so that $\tilde{V}(x, p)$

[4] At the end of the experiment, we also observe an active choice between the two lotteries selected for part 2 between which the subjects should have been indifferent for a subset of our participants. Choice frequencies for lotteries A and B are not significantly different from 50%, and the choice of lottery B is not significantly related to attributes of lottery B. We thus cannot reject the null hypothesis that these participants were indeed indifferent.

[5] Intrinsic utility of this type has been introduced in other domains, such as warm-glow altruism (Andreoni, 1990).

[6] Since $p = 0$ under delegation, monotonicity in $p$ is behaviorally relevant only when $c = 1$.

[7] The indirect utility from delegation, $W(0)$, need not be restricted to expected utility. Formally, delegation induces an Anscombe–Aumann act that maps states—corresponding to the delegate's choice behavior—into objective lotteries over outcomes. The decision maker may evaluate such acts using a preference functional that violates the Anscombe–Aumann independence axiom, thereby allowing for ambiguity aversion (e.g., maxmin expected utility or smooth ambiguity preferences). $W(0)$ can allow for this potentially more general evaluation of delegation, which may incorporate ambiguity-related disutility in addition to expected instrumental payoffs.

corresponds to utility under delegation. This normalization is without loss of generality, as only differences in intrinsic utility between decision modes are behaviorally identified. In particular, $\Delta(x,p,1)$ should be interpreted as the intrinsic utility of retaining the choice relative to delegation. A positive value could thus in principle also reflect a negative intrinsic utility associated with the decision mode $c = 0$ rather than positive intrinsic utility associated with $c = 1$.

The first-stage decision is governed by the comparison of indirect utilities

$$W(1) = \max_{x\in C} \left[\tilde{V}(x,p) + \Delta(x,p,1)\right] \quad \text{and} \quad W(0) = \sum_{x\in C} \pi(x \mid c=0)\, \tilde{V}(x,0),$$

where we use the normalization $\Delta(x,p,0) = 0$.

This representation highlights the general trade-off underlying the decision problem. Retaining the choice ($c = 1$) yields (i) *a potential instrumental benefit of choice*, defined as the possibility to select the most preferred alternative according to $\tilde{V}(x,p)$, and (ii) a potential intrinsic utility component captured by $\Delta(x,p,1)$. Delegation ($c = 0$), by contrast, replaces active choice with a distribution over alternatives given by $\pi(x \mid c = 0)$. Thus, the decision to retain choice autonomy is governed by both instrumental and non-consequentialist considerations.

To make this distinction precise, we define the *instrumental benefit of choosing oneself* as the gain that would arise purely from the ability to select an alternative according to the consequentialist component $\tilde{V}(x,p)$:

$$B^{\text{inst}} \equiv \max_{x\in C} \tilde{V}(x,p) - \sum_{x\in C} \pi(x \mid c=0)\, \tilde{V}(x,0).$$

This object is defined independently of $\Delta(x,p,c)$ and corresponds to the instrumental value of choice under the purely consequentialist benchmark.[8]

For interpretability, we will later consider a reduced-form specification in which the non-consequentialist component is additively separable and independent of $(x,p)$, i.e.,

$$\Delta(x,p,1) = d.$$

Under this reduced-form specification, the overall gain from retaining the choice can be interpreted as reflecting both the instrumental benefit $B^{\text{inst}}$ and the intrinsic utility component $d$.

### 2.2. Purely consequentialist utility benchmark

We now consider the purely consequentialist benchmark as a special case of the general framework, obtained by setting

$$\Delta(x,p,c) = 0 \quad \text{for all } (x,p,c).$$

The first-stage choice then depends only on the trade-off between the gain from selecting $x$ oneself and the cost $p$ associated with it, given the subjective belief about the delegate's choice behavior ($\pi(x \mid c = 0)$). Indirect utilities are then given by $W(1) = \max_{x\in C} \tilde{V}(x,p)$ and $W(0) = \sum_{x\in C} \pi(x \mid c = 0)\, \tilde{V}(x,0)$.

That is, retention of choice is fully governed by instrumental considerations of choice autonomy, and the general trade-off derived above reduces to purely instrumental considerations. In particular, the overall gain from retaining the choice coincides with the instrumental benefit $B^{\text{inst}}$ defined in Section 2.1.

Now, suppose the individual faces a choice set $\tilde{C}$ such that she is indifferent over all alternatives in $\tilde{C}$ at $p = 0$:

$$\tilde{V}(x,0) = \bar{V} \quad \text{for all } x \in \tilde{C}.$$

Then $W(0) = \bar{V}$ and $W(1) = \max_{x\in\tilde{C}} \tilde{V}(x,p)$. Since all alternatives yield the same instrumental utility at $p = 0$, there is no instrumental benefit from being able to choose among them, i.e., $B^{\text{inst}} = 0$. Given that $\tilde{V}(x,p)$ decreases in $p$, pure consequentialism implies indifference between delegation and retaining choice autonomy at $p = 0$, and thus the following behavioral predictions follow[9]:

$$c^* = \begin{cases} 0 & \text{if } p > 0 \quad \text{(delegate)},\\ 1 & \text{if } p < 0 \quad \text{(retain choice)},\\ \{0,1\} & \text{if } p = 0 \quad \text{(indifference)} \end{cases}$$

## 3. The preference elicitation tool

Our empirical approach follows the setup presented in the previous section. To this end, we first deliberately create a choice set $\tilde{C}$ such that the decision maker is revealed to be indifferent between the alternatives contained in the set. Second, we elicit the decision maker's willingness to pay to choose from $\tilde{C}$ herself.

### 3.1. Step 1: Eliciting an indifference set

To create a choice set that only contains alternatives between which the decision maker is revealed indifferent, we made the following design choices: (i) we use choice alternatives whose value can be easily and incrementally adjusted, (ii) we create a simple and easy to understand choice environment to minimize confusion, and (iii) we measure the degree of choice consistency of participants. These design choices are consciously made in order to minimize noise in the process as well as to provide us with usable information about the accuracy with which we have identified the indifference point, given that such identification is crucial for the validity of our measurement tool.

The nature of the alternatives is, in principle, irrelevant. However, because it is important that alternatives can be incrementally adjusted to best approximate an indifference point, we decided to use lotteries over monetary payments as alternatives. Each participant goes through an individual sequence of 10 choice situations in each of which she faces the simple choice between two lotteries $A$ and $B$. Lottery $A$ is fixed and always provides a payoff of 600 points with 25% probability and a payoff of 1600 points with 75% probability. Lottery $B$ provides a payoff of 600 points with 50% probability and a payoff of $X$ selected from $\mathcal{X} \in \{1890,\ldots,2840\}$ points with 50% probability. The value $X$ is adjusted from choice to choice. Probabilities and payments are represented both in numerical and graphical terms.[10]

We adjust $X$ from choice to choice using DOSE—Dynamically Optimized Sequential Experimentation (Wang et al., 2010; Chapman et al., 2022, 2024), meaning that $X$ is always selected in such a way that it maximizes the information gain regarding an individual's indifference point and choice consistency.[11]

Fig. 1 displays three exemplary sequences that help illustrate the procedure and its advantages. The left panel displays the choice sequence and the decisions of a perfectly consistent participant. As one can see, the

[8] Note that, when $\Delta(x,p,c) \neq 0$, the alternative selected under $c = 1$ may differ from the maximizer of $\tilde{V}(x,p)$, so that $B^{\text{inst}}$ does not, in general, coincide with the overall utility difference $W(1) - W(0)$. Rather, it isolates the component of the decision that is driven purely by differences in consequences.

[9] Indifference over all alternatives in $\tilde{C}$ also eliminates ambiguity aversion as a motive for retaining choice autonomy. When $\tilde{V}(x,0) = \bar{V}$ for all $x \in \tilde{C}$, all acts induce identical distributions over utility outcomes. As a result, any preference functional that satisfies monotonicity over Anscombe–Aumann acts—including ambiguity-averse models that violate independence—assigns the same value to all delegation acts. Hence, in the indifference set, ambiguity aversion cannot generate a strict preference against delegation.

[10] A screenshot of the decision screen that participants faced is displayed in Figure A.17 in the Online Appendix.

[11] An exception is choice situations 5 and 10, in which $X$ was chosen via a random procedure that ensured substantial difference of the displayed $X$ value from the values previously seen. We implemented this procedure to break the monotonicity of the choice sequence that would otherwise result for highly consistent subjects.

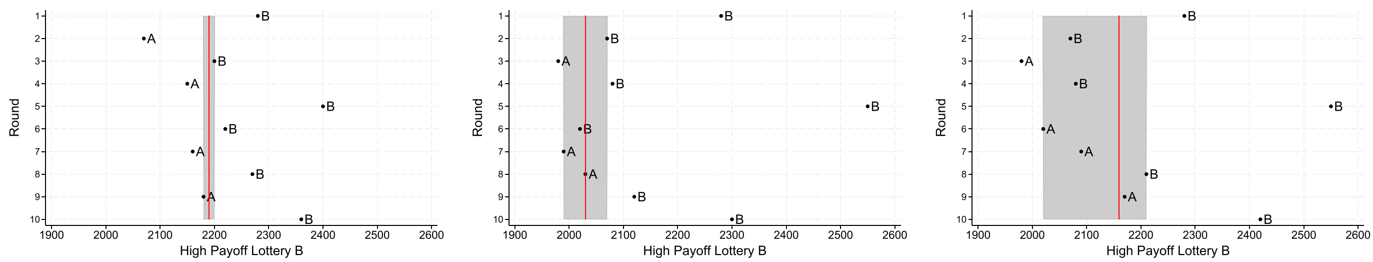



**Fig. 1.** Lottery selection and choice patterns of three exemplary participants. The y-axis of each panel displays rounds 1 to 10. The x-axis indicates the value $X$ of lottery B that was presented in the respective round. Each subject's choice of lotteries A vs. B is shown for each round. Vertical red lines represent the identified indifference lottery. Gray areas illustrate the range outside of which a subject behaved in a perfectly consistent manner. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

DOSE algorithm primarily learns about the indifference point in rounds 1-4, significantly narrowing in on the potential point of indifference. In round 6 the algorithm prioritizes information about the consistency of the subject, by presenting a lottery in which the participant must choose B if she is indeed consistent. The algorithm then proceeds in further narrowing down the indifference point with additional checks on consistency. At the end, the indifference point (indicated by the red line) can be identified within a 10 point range, the smallest unit used in our experimental setup, and the smallest unit expressible in monetary terms, equivalent to less than £0.01.

The middle and right hand panels are illustrative to see how the algorithm enables us to uncover inconsistencies in the choice patterns. The participant in the middle panel makes an inconsistent choice in round 8, while the participant displayed in the right panel makes an inconsistent choice in round 7. As a consequence, DOSE updates its estimate of choice consistency and uses the subsequent rounds to reassess its estimate of the indifference point. The gray areas illustrate the range outside of which a subject behaved in a perfectly consistent manner. Identifying and quantifying such inconsistencies constitutes a valuable feature and significant advancement of our elicitation tool over existing elicitation procedures, because it enables us to control for them in the analysis.

The application of DOSE requires structural assumptions that govern the lottery selection in the background. These assumptions are used solely to implement DOSE in step 1 and do not restrict the general theoretical framework. As the left panel of Fig. 1 makes clear, these structural assumptions also have no impact on the identified indifference point for consistent subjects. They simply govern the sequence of lotteries $B$ that is presented to the subject.[12] In terms of our theoretical framework, alternatives $x \in C$ are instantiated as lotteries over monetary payoffs. We denote expected utility by $U_i(\cdot)$, derived from a Bernoulli utility function over monetary values $u_i(\cdot)$. To operationalize the purely consequentialist benchmark in step 1, we can then assume that $\tilde{V}(x, p)$ is represented by expected utility over monetary outcomes. Since $p = 0$ throughout step 1, utility reduces to $U_i(x)$. Structurally, we assume that an individual's Bernoulli utility function is CRRA, given by

$$u_i(w) = \frac{w^{1-r_i}}{1-r_i} \tag{2}$$

where $w$ is the payoff in points and $r_i$ is the individual's risk aversion parameter. Further, we assume that individual choice behavior is governed by the following probabilistic function

$$Pr_i(A) = \frac{1}{1 + e^{-\mu_i (U_i(A) - U_i(B))}}, \tag{3}$$

where $Pr_i(A)$ is the probability of choosing lottery $A$, and $\mu_i$ specifies individual $i$'s degree of stochastic response in choice. The key individual parameter estimates that we want to obtain are $\hat{r}_i$ and $\hat{\mu}_i$. To this end, we define a discrete parameter space for $r$ and $\mu$. The parameter space for $r$ is determined by the set of payoffs $\mathcal{X}$ of lottery $B$ and contains 96 values given by $\mathcal{R} \in \{-1.2, \ldots, 1.2\}$.[13] $\mu$ can take on the 13 different values given in $\mathcal{M} \in \{1, 10, \ldots, 120\}$, ranging from high randomness in choice to very high consistency in choice.[14] Finally, the function $f(r, \mu) : \mathcal{R} \times \mathcal{M} \rightarrow [0, 1]$ assigns a probability to each parameter combination $(r, \mu)$, and we assume as a prior that the probability distribution over $(r, \mu)$ is uniform. After each choice situation $t$, the joint distribution $f(r, \mu)$ is then updated based on individual $i$'s choice using Bayes' Rule, and based on the posterior distribution, DOSE selects the value $X$ for the next choice situation such that the information gain is maximized.

At the end of step 1, the estimates $\hat{r}_i$ and $\hat{\mu}_i$ are determined based on the posterior joint distribution over $r$ and $\mu$. We can then construct the *indifference lottery* $\hat{B}_i$ that pays a high payoff $\hat{X}_i$ (rounded to the nearest multiple of 10, which is the smallest point unit used in our experiment), such that individual $i$ is expected to be just indifferent between lotteries $A$ and $\hat{B}_i$, thus providing us with a binary indifference choice set $\tilde{C}_i$. Further details about the exact procedures, parameterization, and estimation of $\hat{r}$ and $\hat{\mu}$ are provided in Online Appendix A.1.

### *3.2. Step 2: Eliciting the willingness to pay for choice*

In step 2, participants are then presented with choice set $\tilde{C}$, told that either $A$ or $\hat{B}$ will determine their payoff, and that the choice from $\tilde{C}$ is either made by the participant herself, or by another participant.[15] The two lotteries and the choice between them are represented in exactly the same choice framework as in step 1, simply using $\hat{B}$ as lottery $B$. Subjects thus evaluate the two lotteries in exactly the same decision

[12] In Online Appendix A.1.3, we re-run our estimation of the implied indifference point based on each participant's actual choice data using alternative structural assumptions and show that differences are minimal.

[13] The 96 values contained in the vector $\mathcal{R}$ are determined based on the identifiable parameters of $r$ given the set of lotteries $B$ defined by $X$ above.

[14] Note that $\mu$ is simply a scaling parameter for differences in expected utility (see Eq. (3)). Thus, the values of $\mu$ have to be interpreted in connection with the values of the utility function defined in Eq. (2), and cannot be interpreted in isolation. The value of 0, which implies completely random choice, has been excluded as otherwise DOSE would excessively try to learn whether or not participants' choices are fully random, which hinders learning about the risk parameter given the limited number of choices that we can elicit.

[15] Participants were told that another study participant would choose a lottery on their behalf and receive a fixed base payment. They were informed that the fixed payment is independent of whether the participant delegates and independent of the lottery choices the other participant makes. The other participant made lottery choices for all possible choice situations that may occur in step 2. The data were collected prior to the main experiment in order to guarantee a smooth experience without delay due to matching.

frame in which indifference was elicited, in order to avoid any framing effects affecting the preference over the lotteries.

The key purpose of step 2 is to obtain an estimate of the willingness to pay to choose oneself in this setting. To this end, participants are again faced with a sequence of 10 choice situations, in each of which they must choose between choosing themselves (phrased "I choose") and paying a price $p$, or delegating the choice to an anonymous study participant (phrased "I delegate").[16] The price $p$ can take on values defined in $\mathcal{P} \in \{-600, -590, \dots, -10, 10, 20, \dots, 600\}$, it varies from situation to situation, and it can either be positive or negative.[17] Implementing DOSE for step 2 of our elicitation method again requires putting structural assumptions on $V(x, p, c)$. These structural assumptions are primarily chosen for tractability and computational ease, rather than based on prior beliefs about true structural form. Specifically, we assume that a participant's utility function $V(x, p, c)$ is additively separable in $x$, $p$ and $c$, and can thus be characterized as follows:

$$V_i(x, p, c) = \begin{cases} U_i(A) + d_i - p, & \text{if participant chooses herself } (c = 1) \\ U_i(A), & \text{if participant delegates choice } (c = 0) \end{cases} \tag{4}$$

where $U_i(A)$ is the expected instrumental utility derived from lottery $A$ as defined before (note that $U_i(A) = U_i(\hat{B}_i)$ when $A \sim \hat{B}_i$), $d_i$ captures the difference in intrinsic utility of the participant (as introduced in Section 2.1), and $p$ is the (positive or negative) price she has to pay. Given this utility specification, note that the utility associated with the monetary payoffs of the lottery drops out of the delegation tradeoff, and the price $p$ at which an individual is just indifferent between choosing oneself and delegating is exactly equal to $d_i$.[18] The goal is then to estimate $\hat{d}_i$, and we again apply a DOSE procedure, assuming a probabilistic choice function with consistency parameter $\gamma$ to maximize the information gain about $d$ and $\gamma$ from choice situation to choice situation.[19]

Estimates of $\hat{d}_i$ and $\hat{\gamma}_i$ are then again obtained from the individual posterior distributions over $d$ and $\gamma$. Further details about the exact procedures, parameterization, and estimation are given in Online Appendix A.1.

### 3.3. Implementation and procedures

We conducted two surveys on the platform Prolific Academic (www.prolific.co) in June 2021 and January 2022. Each survey consists of three parts, the two behavioral tasks described in Section 3 and a subsequent questionnaire. Subjects complete part 1 (=step 1) and part 2 (=step 2) as described above. The lotteries are explained in detail before the beginning of part 1 (for illustrations, see Figures A.11 and A.13 in the Online Appendix). For each of parts 1 and 2, one of the 10 choice situations is randomly selected to determine an individual's payment. Feedback on lottery outcomes and payoffs is only given at the very end of the study. The instructions for the behavioral task, including the consent form, can be found in Online Appendix A.3, and the questionnaire is shown in Online Appendix A.4. The questionnaire for the June wave consists of ten blocks of questions that were presented to the participants in random order. We include several scales measuring preferences for and perceptions of autonomy that have been widely used in psychology (locus of control (Rotter, 1966), autonomous functioning (Deci and Ryan, 1985), generalized self-efficacy (Schwarzer et al., 1995) and desirability of control (Burger and Cooper, 1979)) together with several questions capturing personal characteristics, related preferences and socio-demographic information.

**Implementation in June 2021.** The data were collected between June 15th and June 21st 2021 in batches of 200 participants starting at different times of the day. Participants completed the behavioral experiment and could then enter the questionnaire within 24 hours (retention rate of 89.9%). The study was open to anyone,[20] but participants were predominantly Europeans.[21] On average, participants earned £7.09 (consisting of a base payment of £2, an average variable bonus of £2.09, and a payment of £3 for completing the questionnaire) and spent 15.55min on the behavioral experiment and 28.92min on the questionnaire.

**Implementation in January 2022.** The implementation and procedures in January 2022 were similar to June 2021. The replication study was run between January 6 and 12, 2022 on Prolific Academic.[22] We excluded subjects who participated in the previous study and, in addition, we took advantage of the option to restrict participation to subjects with a Prolific Score of at least 99/100, meaning that they behaved in a reliable way in previous studies. The complete study took on average 20.66min and the base payment for participation was £3.5 plus an average variable payment of £2.12.

**Sample composition.** We recruited 998 participants in June 2021 and 794 participants in January 2022. Control questions make sure that participants understand the instructions in the behavioral task and in addition there are several clearly marked attention checks throughout the questionnaire. A participant is excluded from the survey if they fail a set of control questions or the attention checks a third time. The final dataset thus consists of participants who successfully completed the whole study and passed the control questions and attention checks within the allowed number of trials.

From this dataset, we exclude participants who chose lottery $A$ in the choice situation with the highest value of $X$ in their individual sequence, or lottery $B$ in the choice situation with the lowest value of $X$ in their individual sequence. These are individuals who either never switch between lotteries in part 1, and are thus categorized as extremely risk averse/risk loving—and as highly consistent—in our data, or individuals who switch at least once but are inconsistent and display extreme risk aversion / risk lovingness at the boundary of the parameter space. There are different explanations for such choices: subjects might not pay

[16] The experimental interface that participants faced in step 2 is shown in Figure A.27 in the Online Appendix.

[17] Negative prices were framed as "bonuses" that the participant receives if they choose themselves. $p = 0$ was excluded because we expected a significant fraction of participants to have a true willingness to pay of 0, and forcing a choice at 0 (which would reflect a forced tie breaker) would bias our estimate in the positive or negative domain.

[18] We have chosen the additively separable utility specification because of this virtue: If $U_i()$ would not drop out, the DOSE procedure for step 2 would become path dependent on outcomes in step 1, which would significantly increase computational complexity (instead of 1024 potential choice paths in step 1 and step 2 each, the experiment would have consisted of 1024*1024 choice paths, with path dependencies between step 1 and step 2).

[19] In Online Appendix A.1.4, we re-run our estimation of the implied willingness to pay based on each participant's actual choice data using alternative structural assumptions on the utility function. We can show that differences are minimal and estimates are very highly correlated. A discussion of possible micro-foundations of a preference for choice autonomy is provided in Section 6.

[20] Data was collected during the UEFA EURO 2020. Because of the emotions triggered by this event, we excluded nationalities whose teams had a game on the same day.

[21] Nationalities: Greece (5.4%), Italy (6.2%), Mexico (5.7%), Poland (11.2%), Portugal (14.0%), South Africa (9.8%), United Kingdom (13.4%), the remaining nationalities were represented with less than 5% of the sample.

[22] The behavioral task and the (shorter) questionnaire were now run as one study. In addition, the study was run together with another parallel one that addresses different research questions. For this study, we balanced the nationalities that we recruited in the following way: 30–40% Continental Europe, 25% US, 25%, 5–10% South Africa, and 5–10% Mexico. Given these restrictions, we recruited participants with the following nationalities (plus several others with less than 5% of participants in our sample): Italy (7.2%), Mexico (6.4%), Poland (9.9%), Portugal (10.1%), South Africa (6.6%), United Kingdom (23.9%), United States (23.8%).

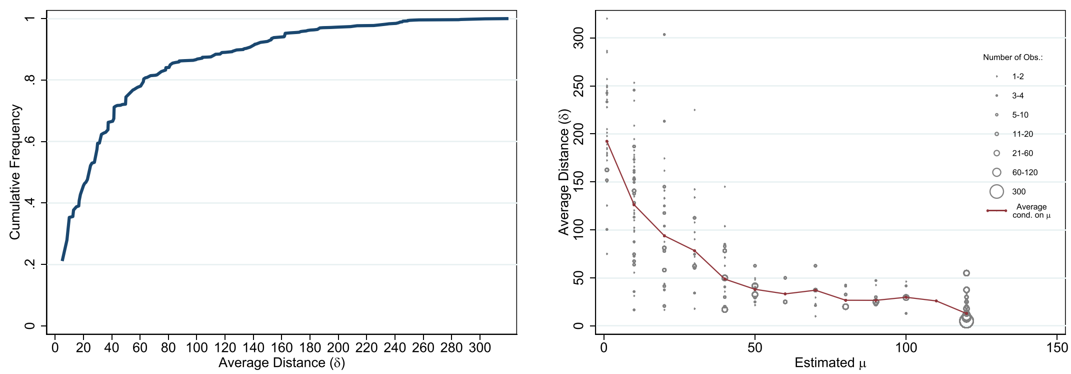



**Fig. 2.** Left panel: Cumulative distribution of $\delta$. Right panel: Scatterplot of $\delta$ conditional on the estimated consistency parameter $\hat{\mu}$.

attention and just click the same button, they might have extreme risk preferences, or they might use a simple heuristic for making the choices between lotteries. While we cannot distinguish these possibilities, the key reason for excluding them is that we cannot identify an indifference set for these participants, and therefore the elicitation of the willingness to pay in part 2 does not work accurately (370 observations). This leaves us with a dataset with 1422 observations.[23]

## 4. Elicitation of indifference in step 1

We begin our analysis by assessing the validity of our identifying assumption: the identification of indifference in step 1 of our elicitation procedure.

We start by quantifying the extent of imprecision in the estimated $\hat{X}$ by determining the range $\rho = [\underline{X}, \overline{X}]$, where $\underline{X}$ is the lowest value of $X$ such that an individual consistently chooses lottery $A$ whenever $X \leq \underline{X}$, and equivalently, $\overline{X}$ is the highest value of $X$ such that an individual consistently chooses lottery $B$ whenever $X \geq \overline{X}$. Thus, $\rho$ identifies the range of potential values $X$ outside of which an individual behaves perfectly consistent, and thus a conservative range for the potential true indifference point. The shaded areas in Fig. 1 display $\rho$ for each of the three exemplary choice patterns. It can be seen that $\rho$ is very small when an individual is consistent, but increases when choice inconsistencies occur.

We can then calculate the average distance between the determined indifference value $\hat{X}$ and the values $X \in \rho$.[24] This average distance is then given by

$$\delta = \int_{\underline{X}}^{\overline{X}} \frac{|X - \hat{X}|}{\overline{X} - \underline{X}} dX \tag{5}$$

Note that the smallest value of $\delta$ in our setting for a perfectly consistent participant is 5, as the smallest unit of variation in $X$ is 10 points.[25] For the participant in the left panel of Fig. 1, $\delta$ is 5 points, it is 20 points for the participant in the middle, and 58 points for the participant on the right.

The left panel of Fig. 2 shows the cumulative distribution of $\delta$ in our data. The average distance is equal to its smallest possible value for 22 percent, and is below or equal to 10 for 36 percent of the sample. The median average distance is equal to 25 points, and 75 percent of the sample have an average distance below or equal to 50 points. Thus, we are able to identify relatively tight ranges of potential indifference points for a large majority of our subjects.

As explained, an advantage of DOSE is that it also directly provides us with a, $\hat{\mu}$. The right panel in Fig. 2 shows how $\hat{\mu}$, the structurally estimated individual choice consistency parameter, relates to $\delta$. It can be seen that for subjects with relatively high degrees of consistency, $\delta$ is small. For subjects with the highest consistency score of $\hat{\mu} = 120$, who constitute 49 percent of our sample, the average $\delta$ is equal to 13.[26] However, even if $\hat{\mu} < 120$, expected deviations remain relatively small as long as subjects do not become too inconsistent. The 28 percent of participants with $\hat{\mu}$ between 40 and 110 on average have a $\delta$ of 36. It is only when $\hat{\mu}$ becomes very small that the precision of the identified point of indifference substantially decreases. In particular, those with $\hat{\mu}$ equal to 1 display substantial distance, implying that the indifference point is very imprecisely estimated.[27] Again, to give an intuition of these values, the participant with the choice pattern in the left panel of Fig. 1 had a $\hat{\mu} = 120$, the participant in the middle had $\hat{\mu} = 80$, and the one on the right had $\hat{\mu} = 10$. Overall, the data reveal that our procedure in step 1 succeeded in identifying points of indifference with substantial precision, and it also succeeded in identifying those participants for whom we cannot ensure that an indifference set was identified with sufficient confidence.

To facilitate this discussion in the remainder of the paper, we will use the structural estimate $\hat{\mu}$ as the individual indicator of choice consistency and define the following subgroups based on the insights we gained in this subsection:

**Definition 1.** A subject is described as highly inconsistent if $\hat{\mu} = 1$, as inconsistent if $1 < \hat{\mu} < 40$, as moderately consistent if $40 \leq \hat{\mu} < 120$, and as highly consistent if $\hat{\mu} = 120$.

4.3% of subjects are thus classified as highly inconsistent, 18.5% as inconsistent, 28.27% are moderately consistent and 49% of subjects are highly consistent.

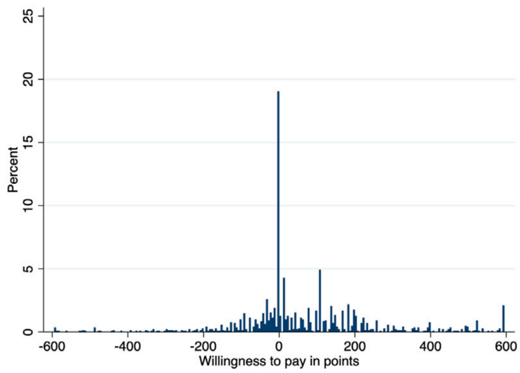
[23] Although identification of indifference is very unlikely, our elicitation algorithm still assigns a value $\hat{B}_i$ based on the posterior distribution of $f(r, \mu)$ to a participant, even if the choice sequence converged to the boundaries of the parameter space (the assigned $\hat{B}$ will then also be close to the boundary of the pre-defined parameter space). For completeness, we reproduce our main analyses using the full sample of 1792 subjects in Online Appendix A.2.3.

[24] Under the assumption that any point on $\rho$ is equally likely to be the true indifference point, $\delta$ would be the expected deviation between $\hat{B}_i$ and the true point of indifference. Note that this is a conservative assumption that likely overestimates the expected deviation, as the selected $\hat{X}$ is based on $\hat{r}$, which is calculated using the posterior probability distribution over $r$ conditional on the actual choice pattern, which is often much less dispersed than a uniform distribution over $\rho$.

[25] To give some context, a variation of 5 points in the high value of the $A$ lottery is equivalent to a 0.28% variation in the expected value of the lottery.

[26] The distribution of the estimated consistency parameters in the June and January samples for both tasks is displayed in Figure A.2 in the Online Appendix. In the elicitation of the willingness to pay in part 2, 61% of subjects are assigned the highest possible estimated consistency parameter given our parameterization, $\hat{\gamma} = 15$. 8.4% of subjects are assigned the lowest possible value, $\hat{\gamma} = 1$.

[27] The heterogeneity in choice consistency partly reflects differences in socio-economic backgrounds. Male participants and participants with a higher education level exhibit somewhat higher consistency scores, especially in part 1 (see Table A.3 in Online Appendix A.2).



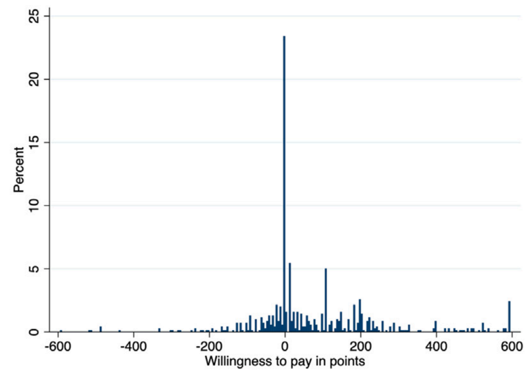




**Fig. 3.** Distribution of the willingness to pay in points (June 2021 and January 2022. LHS: all subjects, $N = 1422$, RHS: only highly consistent subjects, $N = 696$).

Another potential concern is whether the observed willingness to pay for choice autonomy could be driven by preference incompleteness (such as preferences for flexibility (Kreps, 1979; Dekel et al., 2001), or preference adjustment between steps 1 and 2 of our elicitation procedure (even though the experimental design minimizes the scope for preference adjustment by keeping the decision frame identical across steps and varying only a single lottery attribute). Two arguments speak against this concern in our setup. First, participants had no time constraints in contemplating their decision. They could have thus waited for free during the delegation stage in case they were indecisive, instead of paying to retain the choice in order to gain time. Second, among the 743 participants who retained the choice in step 2 and therefore had to actively choose between lotteries $A$ and $\hat{B}$, lottery $A$ was chosen in 49.3% of cases and $\hat{B}$ in 50.7% of cases, frequencies that are not statistically different from 50% ($p = 0.69$). More importantly, if subjects were not truly indifferent, a plausible hypothesis is that participants would systematically favor $\hat{B}$ when payoff $\hat{X}$ is higher (whereas no such pattern should be observable if a higher $\hat{X}$ correctly reflects differences in risk attitudes). In a probit regression of choosing $\hat{B}$ on $\hat{X}$, the estimated coefficient is essentially zero ($\beta = 0.00000321$, $p = 0.99$), and this result also holds when conditioning on participants with a positive willingness to pay in step 2 only. These data thus do not reject the hypothesis that individuals were indeed indifferent over alternatives in $\tilde{C}$.

## 5. Results

### 5.1. Willingness to pay for choice autonomy

The data from our preference elicitation task reveal that individuals on average have a positive willingness to pay for choice autonomy.

**Result 1.**

a) The average and median willingness to pay for autonomous choices in our decision task are significantly positive.
b) The average willingness to pay for choice autonomy amounts to 5% of the total expected value from the choices individuals faced.
c) There is substantial heterogeneity in the willingness to pay for autonomous choices: While the majority of subjects (53.8%) display a positive willingness to pay, it is negative for more than one fourth of subjects (27.1%), and equal to zero for 19.1%.
d) The average willingness to pay for choice autonomy of those who exhibit a positive willingness to pay amounts to 14% of the total expected value from the choices individuals faced. The average willingness to pay for delegation for those who exhibit a negative willingness to pay amounts to 9% of the expected value from the choices.

Support for Result 1(a) is provided in Fig. 3, which shows the distribution of the willingness to pay for all individuals on the left panel, and for the 49% of highly consistent subjects on the right panel. In the whole sample, the average willingness to pay amounts to 67 points, which is significantly different from 0 ($t$-test: $p < 0.001$, N= 1422). The median willingness to pay is 12 points (Wilcoxon sign-rank test (in the following WSR): $p < 0.001$, N= 1422). When conditioning on being highly consistent in part 1, the average willingness to pay amounts to 76 points and the median remains at 12 points ($t$-test, WSR: $p < 0.001$, N= 696).[28]

To support Result 1(b), we calculate how much individuals are willing to give up for choice autonomy in terms of the value of the underlying decision at hand. Expressed in expected value, we find that the willingness to pay for choice autonomy on average amounts to 5% of the expected value of lottery $A$ (5.6% for the highly consistent, 4.9% for those with a very small $\delta \leq 10$).

Result 1c) states that there is substantial heterogeneity in the measured preferences. 53.8% of individuals have a strictly positive willingness to pay, 19.1% have a willingness to pay of zero, and for 27.1%, the willingness to pay is strictly negative. This implies that preferences over autonomous choices can take a positive or a negative value, the latter expressing an aversion to choosing for oneself. The distribution again looks fairly similar when conditioning on the subsample of highly consistent participants (54.74%, 23.4%, and 21.8%).

In Result 1(a), positive and negative willingness to pay is averaged, netting out expressed preferences for and against choice autonomy, even though these quite likely constitute succinct preference types. Another way of looking at the data is that both express a positive willingness to pay for a specific decision mode. Conditional on expressing a positive willingness to pay (indicating preferences for choice autonomy), the average willingness to pay to retain choice amounts to 184 points (N= 765, SD=161, median=136 points), or 14% of the expected value of the choice they faced. Conditional on expressing a negative willingness to pay (indicating preferences against choice autonomy), the average willingness to pay to delegate amounts to 119 points (N= 386, SD=117, median=66 points), which constitutes a share of 9% of the expected value of the choice (Result 1d)).

Finally, it is noteworthy that some individuals indicate either a very high or very low willingness to pay for autonomy (see Fig. 3), and one might be worried that these values do not reflect the true preferences. Our results are robust to excluding participants who indicate an extremely high or extremely low willingness to pay (i.e., those who always accepted or always rejected an offer in part 2 and thus exhibit a WTP at one of the corners of the range that we can estimate). The average WTP remains positive without these observations and it amounts to 57.9 points, which is still significantly different from 0 ($t$-test: $p < 0.001$, WSR: $p < 0.001$, N=1387).[29]

[28] Similarly, the average willingness to pay of the 36% who exhibit a very small average distance of $\delta \leq 10$ is 66 points ($t$-test, WSR: $p < 0.001$, N= 502).

[29] 30 subjects were willing to pay every price offered to them. 5 subjects were willing to delegate for every bonus offered to them. To ensure that our results

To explore heterogeneity in intrinsic preferences for choice autonomy and provide a first assessment of potential antecedents, Table A.5 in Online Appendix A.2 shows correlations with socio-demographic characteristics and personal attitudes, and shows that gender is significantly associated with preferences for choice autonomy. Male participants on average exhibit a higher willingness to pay for choice autonomy. All other socio-demographics and other characteristics are of limited explanatory power. In particular, reported trust in others (see Online Appendix A.4 for exact questions) is not related to the willingness to pay, which supports our previous conclusion that we successfully excluded instrumental benefits of choice (i.e., utility gains from selecting the better option).

Only a person's reported willingness to take more risks is correlated with willingness to pay: the more willing to take risks, the higher the willingness to pay. This suggests that an intrinsic preference for autonomy and a willingness to take more risks may be related characteristics (see also Dean and Ortoleva (2019) for a discussion of related behavioral characteristics).[30]

To assess whether our results are influenced by the online environment in which the study was conducted, we replicated the main results in May 2022 in a lab experiment at the University of Fribourg, Switzerland. We recruited 152 participants for a total of 13 sessions. In each session, one participant was randomly selected to be the person to whom all other participants could choose to delegate their choice. Consequently, 139 subjects were decision makers in part 2 of the experiment, 122 of whom passed the exclusion criteria outlined in Section 3.3.[31] Stakes in the lab setting were significantly higher, although the framing of the experiment in terms of points remained the same.[32]

In the lab sample, the average willingness to pay to choose oneself is 43.3 points (the median willingness to pay is zero), which is again significantly different from 0 ($t$-test, WSR: $p < 0.001$, N = 123). Overall, 31.7% of the students display a willingness to pay of zero, 22.8% have a negative willingness to pay and 45.5% a positive willingness to pay. Those with a positive willingness to pay have an average of 140 points (median=120), while the average value for those with a negative willingness to pay amounts to 89 points (median=62).[33] The average and the median values in terms of points are thus somewhat smaller in magnitude compared to the online sample, and we observe fewer participants with a positive willingness to pay and more participants with zero willingness to pay for choice autonomy. Overall, the results from the in-person lab experiment confirm the results obtained from the Prolific sample (the differences in mean and median willingness to pay are not significant, $t$-test/MWU with p>0.1), suggesting that the results are not an artifact of the online environment on Prolific, or the relatively small stakes involved in a large scale online experiment.

### 5.2. Choice inconsistency and instrumental value

Our analysis in Section 5.1 shows that participants on average have a positive willingness to pay to retain choice autonomy, and reveals that this result holds for subjects with precisely identified indifference sets. In this subsection, we further explore the relationship between our preference measure and the precision with which the indifference set is identified.[34]

**Result 2.** The expected instrumental benefit of choice is negligible except for highly inconsistent subjects. For all subjects, we observe that the instrumental benefit of choice due to imprecise measurement of the indifference set cannot explain the measured willingness to pay for choice autonomy.

The analyses leading to Result 2 are displayed in Fig. 4. The top left panel of Fig. 4 displays the average willingness to pay to retain choice autonomy for the four consistency classes defined in Section 4. It can be seen that the average willingness to pay is significantly positive for all four categories. Consistent with the notion that instrumental value differences between $A$ and $\hat{B}$ may exist when $\tilde{C}$ is estimated less precisely, and thus retaining choice autonomy may generate instrumental benefits of choice, we do observe a decrease in the willingness to pay when moving from inconsistent to consistent participants, although the overall trend is not very strong and not statistically significant. However, perhaps surprisingly, the average willingness to pay is lowest for the subgroup with $40 \leq \hat{\mu} \leq 120$. Indeed, the willingness to pay of this subgroup is (weakly) significantly smaller than the willingness to pay of the highly inconsistent subgroup ($t$-test: $p = 0.059$) and significantly smaller than the willingness to pay of the highly consistent subgroup ($t$-test: $p = 0.006$). The difference to the inconsistent subgroup is not statistically significant.

Furthermore, the distribution of the willingness to pay between the consistent and highly consistent subgroups differs significantly. Among the highly consistent subjects, 22% display a negative willingness to pay, 23% a willingness to pay of zero, and 55% a positive willingness to pay. In contrast, the willingness to pay is negative for 34% of the moderately consistent subjects, zero for 15%, and positive for 51%. These distributional differences are statistically significant ($p < 0.001$, Kolmogorov–Smirnov test, N=402+696).

To further deepen our understanding of the potential relevance of inconsistent choice on the measured willingness to pay, we next consider whether instrumental benefits from choice might explain the observed willingness to pay for choice autonomy in step 2 of our elicitation process. To this end, we utilize the CRRA utility specification introduced in Section 3. In particular, we use the posterior probability distribution $f(r|\hat{\mu})$ (see Section 3.1 and Online Appendix A.1 for details) to estimate the expected *instrumental benefit of retaining choice autonomy*, which is given by the difference in instrumental value between options $A$ and $\hat{B}$, given the risk preference parameter $r$. If a participant is presented with an indifference lottery $\hat{B}_i$ based on $\hat{r}$, but the true risk preference parameter is $\tilde{r}$, the instrumental benefit of choice is given by:

$$\text{Instrumental Benefit of Choice: } B_{choice} = |U(A|\tilde{r}_i) - U(\hat{B}_i(\hat{r}_i)|\tilde{r}_i)|,$$

where $U(A|\tilde{r}_i)$ is the expected instrumental utility of the fixed lottery A, and $U(\hat{B}_i(\hat{r}_i)|\tilde{r}_i)$ is the expected instrumental utility of the lottery $\hat{B}_i$

are not driven by inconsistent subjects, we also replicate all results excluding subjects with $\hat{\mu} < 40$ in Online Appendix A.2.2.

30 Note that, if anything, delegation increases outcome risk. Thus, this observation is inconsistent with any interpretation related to instrumental value—if instrumental value would drive this correlation, a person should be less willing to delegate if she is risk averse. While one might worry that this correlation contains some experimental confound since we elicit the individual's willingness to pay via choices between lotteries, note that columns (3-6) show that the correlation is equally strong between our measure of risk taking and independent survey measures of a preference for or perception of free choice: we observe a significant positive correlation with the World Value Survey question on freedom and control (Inglehart et al., 2014) and with the Desirability of Control Index (Burger and Cooper, 1979). The correlation is mostly—but not perfectly—robust to the inclusion of control variables and to the estimation method, see Table A.7 in the Online Appendix.

31 Subjects were university students. The person to whom decisions were delegated participated normally in step 1, and received a fixed payment of CHF 10 for part 2. The experimental design and presentation was otherwise identical to the studies on Prolific. The payoffs were adapted to commonly paid salaries for student jobs in Switzerland (CHF 24.25 on average, including a CHF 10 show-up fee).

32 Higher stakes were achieved through a higher exchange rate in the laboratory, where 1000 points were worth CHF 5.

33 In USD, these amount to 0.91 USD and 0.6 USD, respectively.

34 We also measure choice consistency in part 2. The distribution of the modal consistency parameters in both parts is shown in Figure A.2 in the Online Appendix. Choice consistency is higher in part 2 with 62% highly consistent subjects. Additionally, 68% of those who are highly consistent in part 1 are also highly consistent in the second part. The correlation of both consistency scores is highly significant but rather moderate with a correlation coefficient of.12.

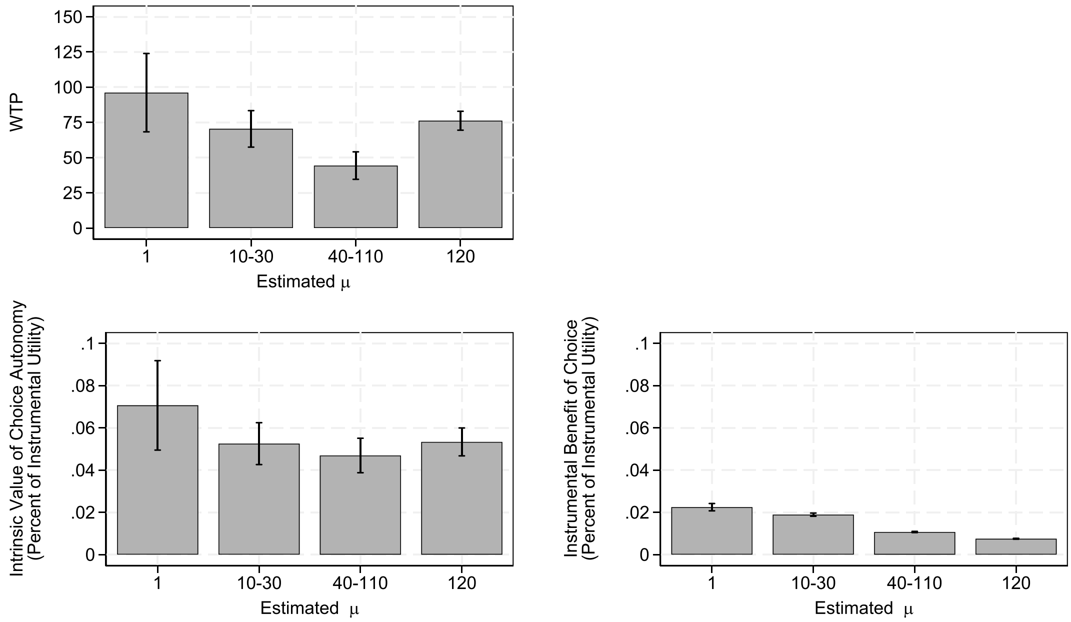



**Fig. 4.** Top left panel: relationship between the estimated mean willingness to pay and the estimated consistency parameter $\hat{\mu}$. Bottom left panel: relationship between the expected average intrinsic value of choice autonomy (see Online Appendix A.1.5 for details on the calculation), expressed in percent of the expected utility of the underlying lottery $A$, and the estimated choice consistency parameter $\hat{\mu}$. Panel 3: relationship between the expected instrumental benefit from choice, expressed in percent of the expected utility of the underlying lottery $A$ (calculated as explained in Eq. (6)), and the estimated choice consistency parameter $\hat{\mu}$.

that is chosen based on the estimated risk preference parameter $\hat{r}_i$, given the true risk preference parameter $\tilde{r}_i$. Thus, for each participant, we can calculate the expected instrumental benefit of choice given $\hat{B}_i(\hat{r}_i)$ and given $f(r|\hat{\mu}_i)$:

$$E(B_{choice}) = \int_{\mathcal{R}} f(r|\hat{\mu}_i)|U(A|r) - U(\hat{B}_i(\hat{r}_i)|r)|dr \tag{6}$$

The bottom right panel of Fig. 4 displays the relationship between the expected instrumental benefit of choice and the estimated consistency parameter $\hat{\mu}_i$, expressed as a percentage of the utility obtained from lottery $A$ (assuming that $\hat{r}_i = \tilde{r}_i$). We see that the expected instrumental benefit of choice is substantial when subjects are highly inconsistent, but decreases as consistency increases. The mean expected instrumental benefit of choice drops to 1 percent of $U(A|\tilde{r}_i)$ once subjects are at least moderately consistent ($\hat{\mu}_i \geq 40$); see bottom right panel of Fig. 4. The bottom left panel of Fig. 4, in contrast, shows the average estimated intrinsic value of choice autonomy for all levels of $\hat{\mu}_i$, also expressed as a percentage of $U(A|\tilde{r}_i)$.[35]

The difference between the measured intrinsic value of choice autonomy and the calculated expected instrumental benefit of choice is significantly different from zero using a $t$-test, both overall ($p < 0.01$) and for each of the $\hat{\mu}$ categories shown in Fig. 4 ($p = 0.026$ for $\hat{\mu}_i = 1$, $p < 0.001$ for $\hat{\mu}_i > 1$). Thus, an instrumental benefit of choice due to imprecise measurement of the point of indifference in step 1 cannot explain the measured willingness to pay for choice autonomy in our experiment.

Combining the insights from all panels of Fig. 4 as well as our analyses in Section 4 suggests that it is only for highly inconsistent subjects that we find a possibly substantially inflated willingness to pay because of a substantial instrumental benefit of choice. However, for subjects with moderate to high levels of consistency ($\hat{\mu}_i > 30$) inconsistencies are fairly small and any remaining instrumental benefit of choice does not seem to play a major role in determining the individual willingness to pay.

### 5.3. *Delegating to a coin toss*

In order to better understand the potential determinants of the observed behavioral preference for choice autonomy in our decision task, we conducted a control experiment in which participants had the possibility to delegate the choice between the two lotteries to a coin toss that chooses either option with 50% probability in step 2 of the elicitation procedure (treatment COIN). All other procedures and design features remained identical to the previously described experiment (treatment HUMAN).

Whether or not delegating to a coin constitutes a loss of choice autonomy remains an open question, and probably depends on individual perception. On the one hand, one argument is that the decision maker fully retains choice autonomy even in case of delegation because delegation simply constitutes the deliberate choice of an objective compound lottery over the two individual lotteries, expressing the deliberate preference over lotteries of the decision maker. On the other hand, it remains conceivable that delegation continues to constitute a loss of choice autonomy also in the COIN treatment, as the decision maker forgoes the possibility to choose a specific alternative and thus loses agency in determining her outcomes.[36]

We collected data from 246 subjects on August 22, 2023 on the platform Prolific.co.[37] Applying the same exclusion restrictions as before,

[35] Calculating the intrinsic value of choice autonomy as a percentage of the instrumental utility of lottery $A$ requires structural assumptions on $V(x, p, c)$. Here, we maintain the assumption used in the implementation of DOSE in step 2 that the intrinsic value of choice autonomy is additively separable from the instrumental utility, but, diverging from the previous structural implementation, now assume that $p$ is integrated with the lottery outcomes and evaluated using CRRA utility. If $p$ remained additively separable, $U(A)$ would drop out of the utility comparison between $c = 1$ and $c = 0$, implying that any percentage measure relating the intrinsic value of choice autonomy to $U(A)$ would be ill-defined. Details on the exact calculations and structural assumptions are provided in Online Appendix A.1.5.

[36] In principle, there is no a priori reason to believe that individuals couldn't even be *more* averse when delegating to an automated draw (which they may perceive as a "machine") than when delegating to a human decision maker.

[37] US Americans were not allowed to enter the study because we collected data with only US American subjects for another (large) experiment at the same time.

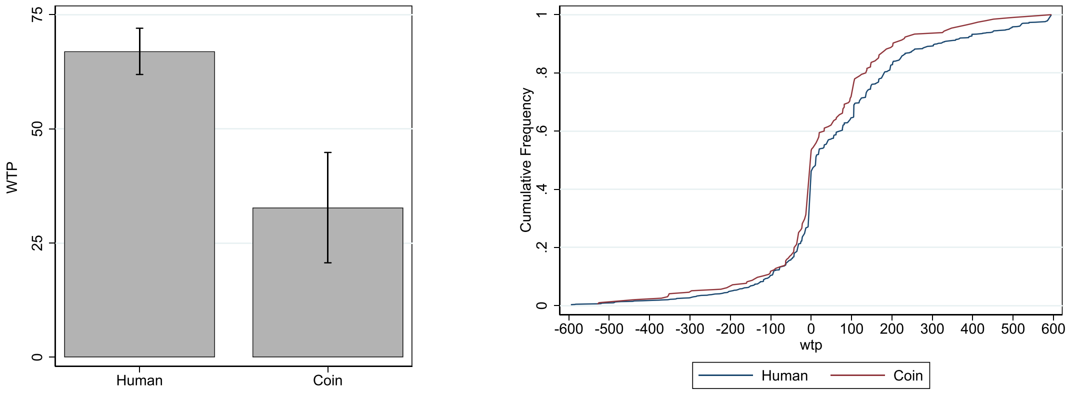



**Fig. 5.** Left Panel: Average willingness to pay in points by treatment, N=1422 (human) and N=195 (coin). Right Panel: Cumulative distribution of the willingness to pay by treatment.

we are left with 195 observations. We find that the willingness to pay in COIN is also positive: The mean value is 32.75 (std. dev.=168.66), which is still significantly larger than zero ($t$-test, WSR: $p < 0.01$, N= 195), but half the magnitude than in HUMAN. This difference is statistically significant (two-sample $t$test p=0.018, MWU: p=0.032). Among the 97 highly consistent subjects, the mean willingness to pay is 14.82 points, which is no longer significantly larger than zero. The individual heterogeneity shows a similar pattern: The willingness to pay is positive for 46.67%, zero for 22.05%, and negative for 31.28%. Thus, the share of subjects with a positive willingness to pay is somewhat smaller compared to the HUMAN treatment. The mean willingness to pay conditional on a positive willingness to pay amounts to 152.49 points (median=105.82) and to 122.78 points (median=60.57) conditional on a negative willingness to pay.

Fig. 5 contrasts the findings from the COIN and HUMAN treatments. The left panel shows that the average willingness to pay in COIN is indeed significantly smaller ($t$-test: $p = 0.02$; MWU: $p = 0.03$). Restricting the sample to those who are at least moderately consistent or highly consistent confirms these results. The right hand panel shows the distribution of the willingness to pay to retain choice autonomy across the two experiments. It appears that there is a shift particularly the willingness to pay towards smaller values in in the positive domain, although distributional differences are only marginally significant (Kolmogorov-Smirnov Test: $p = 0.09$). Again, restricting the sample to those participants who are moderately or highly consistent confirms these results ($p = 0.08$ and $p = 0.07$, respectively).

**Result 3.** When subjects can delegate to an objective random draw, the mean willingness to pay to choose oneself is significantly positive, but significantly smaller compared to the situation in which subjects can delegate to another human decision maker (33 compared to 67 points).

The results of the COIN treatment reveal that the displayed preference to retain choice autonomy becomes significantly weaker when choice is delegated to an objective coin toss rather than a human decision maker. It thus appears that the existence of a human decision maker, or more generally the identity of the delegate, is a fundamental determinant of the value of choice autonomy. While it is not obvious whether (all) individuals perceive delegation to an objective coin toss as a loss of choice autonomy, the fact that we still observe a significantly positive willingness to pay suggests that there is substantial heterogeneity in this perception.

### 5.4. Related psychological constructs

How does our measured preference relate to established concepts in psychology? Our questionnaire includes measurements of locus of control (LOC, Rotter (1966)), autonomous functioning (IA, Deci and Ryan (1985)), generalized self-efficacy (GSE, Schwarzer et al. (1995)) and desirability of control (DC, Burger and Cooper (1979)), as well as the World Value Survey question measuring a person's perceived freedom and control (Inglehart et al., 2014).[38]

These measures are among the most widely-used and highly established measures for autonomy and control in the psychology literature. Because of their wide and frequent use, we consider it important to explore how our novel measure relates to them. Our measure of an intrinsic value of choice autonomy singles out and identifies one psychological phenomenon that is closely related to, or part of, the concepts of control, self-efficacy and autonomy captured in the survey measures. Note that there is nevertheless no prior hypothesis as to whether to expect a correlation: While our measure elicits a *preference*, the four survey measures express a *perception*, in other words, a person's belief about the degree of autonomy (freedom/control/self-efficacy) that she has. Verme (2009) argues that for a person to value freedom of choice she has to believe to have an internal locus of control—because the latter allows her to take advantage of free choices. On the other hand, experiencing restricted freedom of choice may be associated with a low locus of control, and could induce a strong desire for autonomy in the individual. Since there is no prior data to substantiate a hypothesis, we explore possible associations.

Table 1 displays OLS regressions of the psychological constructs on WTP with standardized variables, in column (1) without and in column (2) with controls.[39] The correlations are generally low and insignificant.[40]

(Chapman et al., 2025) suggests that correlations between self-assessed and incentivized measures of a preference or attitude may generally be weak. In our case, the variables we employ also measure theoretically different underlying behavioral dispositions. The lack of correlations across all five survey items indicates that an intrinsic preference for choice autonomy may be conceptually distinct from and

Except for this restriction and an approval rate of at least 99/100, there were no additional pre-screening criteria. The average payoff was £4.15, including a £2 base payment. The median duration was 18min.

[38] For each measure, we use the original questions and answer scales; see questionnaire items in Online Appendix A.4.

[39] In Tables A.8 and A.9 in the Online Appendix, the same estimations are repeated using OLS regressions without standardized variables as well as median regressions.

[40] In contrast, the correlations between the psychological constructs themselves are large and significant (see A.10 in Online Appendix A.2), confirming previous literature (Aldama et al., 2021).

**Table 1**
Each cell shows the coefficient of one OLS regression (with robust standard errors) with z-standardized covariates and z-standardized willingness to pay as the dependent variable. Constants are omitted. Independent variables in the 10 regressions are: LOC: index of internal control (Rotter, 1966), IA: index of autonomy (Deci and Ryan, 2006), GSE: self-efficacy index (Schwarzer et al., 1995), DC: index of desirability of control (Burger and Cooper, 1979), WVS: World Value Survey question on freedom and control (Inglehart et al., 2014). The five regressions in column (2) include controls for age, gender, income, education, risk_taking, nationality, highly_inconsistent_part1, highly_inconsistent_part2, prolific_score and not_failed, where binary/categorical variables have not been standardized. The data is from the June 2021 wave.

| | WTP | |
|---|---|---|
| LOC | 0.029 | 0.005 |
| | (0.038) | (0.040) |
| IA | 0.008 | 0.012 |
| | (0.036) | (0.039) |
| GSE | 0.050 | 0.025 |
| | (0.036) | (0.042) |
| DC | 0.057 | 0.030 |
| | (0.033) | (0.037) |
| WVS | 0.062 | 0.046 |
| | (0.038) | (0.040) |
| Controls | no | yes |
| Observations | 791 | 782 |

independent of the perceptions captured by the traditionally used survey items.

**Result 4.** The willingness to pay for choice autonomy in our task is not related to well-established survey-based measures used in the psychology literature. This indicates that we measure a psychological construct that is different from perceptions of one's own autonomy, locus of control and self-efficacy.

## 6. Discussion

Our preference elicitation method reveals that individuals have intrinsic preferences for choice autonomy and shows how they differ with respect to this preference. Comparing our results to findings of studies that looked at potentially related constructs of intrinsic utility of control, we observe that the willingness to pay for choice autonomy is lower than that for decision rights in Bartling et al. (2014) or the control premium measured in Owens et al. (2014). This is consistent with the fact that our measure more precisely isolates a single motive, choice autonomy, and more effectively eliminates potential confounding factors, such as ambiguity or instrumental benefits of choice due to imprecise measurement. The intrinsic value of choice autonomy measured in our experiment can have multiple microfoundations, several candidates of which are discussed below.

**Autonomy is a direct source of utility.** One perspective is that autonomy directly functions as a consumption good ($c$ enters as an additively separable component in $V(x, p, c)$). This view is consistent with the idea that autonomy is a human need (Deci and Ryan, 1985) and can generate procedural utility (Frey et al., 2004). Such a need could either constitute positive consumption value from exercising autonomy, or it could stem from an aversion to being other-determined. It could also be microfounded in self-image concerns (Bénabou and Tirole, 2006), whereby retaining decision authority reinforces a positive sense of agency, competence, or related aspects of the self. The fact that subjects facing the possibility to delegate to a random draw display a smaller average value for choice autonomy suggests that an aversion to having choices over one's own payoff consequences made by *another person* can partly explain the observed value for autonomous choice. If one adopts the view of choice autonomy as having consumption value, it becomes pertinent to explore the structure of the respective utility function and its manifestations. For instance, if autonomy resembles a consumption good, one may expect that it displays decreasing marginal utility. For our elicitation tool, it would imply that it measures the marginal utility received from an additional choice, which may depend on the degree of choice autonomy that a decision-maker experiences in everyday life.

**The process of choice changes the utility of outcomes.** Another possibility is that agency in choice changes the utility that the decision maker receives from specific outcomes ($c$ interacts with $x$ in $V(x, p, c)$). For example, the (non)-involvement in the choice process may affect the feelings and emotions that are experienced when specific outcomes are obtained, and thus the process of choice may affect the utility associated with the lotteries contained in the choice set in our experiment. Such a specification would also imply that the intrinsic value of choice autonomy heavily depends on $x$, potentially scaling with the instrumental utility of $x$ or specific attributes of $x$.

Taking such a specification one step further, the utility function could also be defined over the metachoice (Bernheim and Taubinsky, 2018). Step 1 of our elicitation task only reveals indifference between lotteries $A$ and $B$ in the decision frame in which the lotteries are evaluated without the option to delegate. A positive willingness to pay in step 2 then reveals that the decision maker has a preference for choice autonomy in the frame in which she faces the possibility to delegate that choice to another individual (for example, due to feelings and emotions associated with the two procedures of choice). An important further implication of this view is that, as Bernheim and Taubinsky (2018) point out, one faces the so-called comparability problem. Our decision experiment only reveals a preference for choice autonomy in a situation in which the decision maker faces a choice to delegate. While we believe that this is a highly relevant and pertinent decision context, it is in principle possible that the decision maker would be generally better off in a situation in which she never faced such a choice, and another party simply chose lottery $A$ or $B$ for her. However, the utility associated with such a decision frame remains unobserved.

**Preferences for/against randomization.** Finally, as discussed in the introduction, another potential partial explanation of the observed delegation and non-delegation behavior can be found in models of deliberate stochasticity (Machina, 1985; Cerreia-Vioglio et al., 2019). The primary application of these models so far has been the explanation of expressed preferences for randomization (Agranov and Ortoleva, 2017, 2023) via convex preferences over lotteries, a close equivalent of which would be delegation in our setting, and which could be expressed via a *positive willingness to pay to delegate*. In fact, the share of subjects with a negative willingness to pay in our data (27.1% in HUMAN, 31.28% in COIN) is quite similar to the 29% who are willing to pay to delegate in Agranov and Ortoleva (2017) (but less than the 40% found in Agranov and Ortoleva (2023)). However, this literature notes that randomization preferences could also go the other way. For example, Machina (1985) has argued that an individual could also be "(weakly or strictly) averse to randomization over indifferent alternatives" (Machina, 1985, p.590), and Cerreia-Vioglio et al. (2019) write that "the model […] also permits indifference to randomization, […] or even aversion to randomization" (Cerreia-Vioglio et al., 2019, p.2430). Such behavior would imply concavity in preferences over lotteries.

The COIN treatment in particular offers an experimental setup that may inform the models of Cerreia-Vioglio et al. (2019) and Machina (1985) with respect to the potential relevance of concavity in addition to convexity in preferences over lotteries, given that randomization is objective, and an indifference choice set was deliberately constructed. Our data suggest that aversion to randomization is in fact more frequent than preferences for randomization.[41]

Preferences for/aversion towards randomization rather than intrinsic preferences for choice autonomy thus potentially provide an alternative explanation for the observed behavior in our experiment. But can it explain the whole picture? As argued, whether or not delegating to a coin can still be thought of as a loss of choice autonomy remains an open question, and probably depends on individual perception. Any motives for/against randomization stemming from preferences for randomization, however, remain present. The fact that the willingness to pay to retain choice autonomy is significantly larger in the HUMAN treatment compared to the COIN treatment suggests that considerations of agency and self-determination, as well as a potential aversion to being dependent on the choice of another human decision-maker (see also Freundt and Herz (2024)), are thus additional factors that influence the expressed preference for choice autonomy and that point towards concerns for autonomy.[42]

One aspect of our data that is, however, consistent with preferences for randomization, is the fact that moderately consistent subjects have the lowest average willingness to pay and the highest fraction of a negative willingness to pay in our data. Agranov and Ortoleva (2023) show that preferences for randomization manifest themselves around a potential indifference point, and do not imply randomness over the whole interval. In our setting, such preferences would thus indeed most likely be classified as being moderately consistent, but not highly consistent or inconsistent. At the same time, preferences for randomization would predict that subjects display a *negative* willingness to pay when prompted with a choice from a set that contains alternatives between which the participant is (nearly) indifferent. In contrast, highly consistent subjects do not display preferences for randomization in step 1, and are thus also less likely to display a negative willingness to pay for choice autonomy in step 2 of our elicitation task.[43]

The role of intrinsic preferences for choice autonomy and randomization preferences in explaining (non-)delegation behavior is an interesting direction for future research.

## 7. Conclusion

In this paper, we develop a novel incentivized behavioral measurement tool for intrinsic preferences for choice autonomy that is suitable for wide-scale applications in the laboratory as well as in online and lab-in-the-field experiments. Our innovative elicitation method identifies the intrinsic utility component of choice autonomy while excluding any instrumental benefit that is attached to choice and controlling for measurement error. We provide evidence for the existence of such preferences in a large online sample and conclude that the value of choice autonomy cannot be reduced to its purely instrumental benefits. We also find substantial heterogeneity in preferences, with about half of the subjects exhibiting a positive value, but a sizeable minority of more than a fourth of participants showing an aversion to making autonomous choices.

We find that the average willingness to pay is reduced to about half the size in another experiment in which participants faced the choice to delegate to a random draw instead of to another person, suggesting that it matters to whom one can delegate. Future research is needed to better understand how the identity of the delegate shapes preferences for choice autonomy.

Our approach has so far only considered choices over lotteries. How such preferences generalize over domains and to a wider range of economic and political behavior remains an open question. It seems conceivable that intrinsic preferences for autonomy play an important role in these settings. For example, Sugden et al. (2008) or Arad and Rubinstein (2018) suggest that a preference for freedom of choice might matter in how some people react to libertarian paternalist policies, including commonly employed nudges such as defaults. Similarly, one cause of the British vote to leave the EU appears to have been a desire to “take back control” (May, 2017).

Following the recent experimental literature on paternalism, our methodology may also be valuable for better understanding how people make decisions for others. Ambuehl et al. (2021) suggest that people may project their own preferences onto others when they are able to influence their choice sets. Such preferences for making choices for others are conceptually distinct, but may be correlated with a preference for choice autonomy. In addition, Ackfeld and Ockenfels (2021) suggest that people tend to take the other person’s autonomy into account when engaging in paternalistic actions. It is thus of interest to the literature on paternalism to understand if and when people exhibit an intrinsic preference for making their own choices.

Intrinsic preferences for choice autonomy may also have an important impact on optimal organizational design. Dessein and Holden (2022) show theoretically how private benefits stemming directly from decision making, such as intrinsic preferences for choice autonomy, can shape organizational design. Bloom et al. (2012) show a cross-country correlation between the degree of decentralization of organizations and the power distance index (Hofstede, 2001), which “measures the perceptions of and preferences for hierarchical relationships” (Bloom et al. (2012), p.1687). Differences in decentralization can be caused by differences in intrinsic preferences for autonomy, but are likely also caused by differences in the instrumental value of different organizational forms across countries (see also Freundt and Herz, 2025). Our tool can help in investigating the factors underlying such relationships.

## CRediT authorship contribution statement

**Jana Freundt:** Writing – review & editing, Conceptualization, Validation. **Holger Herz:** Writing – review & editing, Conceptualization, Validation. **Leander Kopp:** Software, Conceptualization.

## Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

Holger Herz, Jana Freundt and Leander Kopp report that financial support was provided by European Research Council. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Appendix A. Supplementary data

Supplementary data to this article can be found online at doi:10.1016/j.jpubeco.2026.105681.

## Data availability

Data will be made available on request.

[41] A counterargument against interpreting a willingness to pay to delegate in our setup is that participants could have randomized in their heads without a need to pay. However, this argument relies on the assumption that humans are capable of performing such randomization.

[42] Obviously, the comparison between the two treatments is not perfect, given that the COIN treatment provided objective probabilities, whereas choice probabilities in HUMAN remained uncertain. However, we would argue that such uncertainty is inherently linked to a loss of autonomy.

[43] Note that another potential explanation for the increase in the willingness to pay when moving from moderate consistency to full consistency could be that those who intrinsically value choice are more likely to take the choice seriously (because they value it more) and are thus more likely to have consistent choice patterns.

## References


Ackfeld, V., Ockenfels, A., 2021. Do people intervene to make others behave prosocially? Games Econ. Behav. 128, 58–72.

Agranov, M., Ortoleva, P., 2017. Stochastic choice and preferences for randomization. J. Polit. Econ. 125 (1), 40–68.

Agranov, M., Ortoleva, P., 2023. Ranges of randomization. Rev. Econ. Stat. 1–44.

Aldama, A., Bicchieri, C., Freundt, J., Mellers, B., Peters, E., 2021. How perceptions of autonomy relate to beliefs about inequality and fairness. PLOS ONE 16 (1).

Ambuehl, S., Bernheim, B.D., Ockenfels, A., 2021. What motivates paternalism? An experimental study. Am. Econ. Rev. 111 (3), 787–830.

Andreoni, J., 1990. Impure altruism and donations to public goods: a theory of warm-glow giving. The economic journal 100 (401), 464–477.

Arad, A., Rubinstein, A., 2018. The people's perspective on libertarian-paternalistic policies. J. Law Econ. 61 (2), 311–333.

Astebro, T., Herz, H., Nanda, R., Weber, R.A., 2014. Seeking the roots of entrepreneurship: insights from behavioral economics. J. Econ. Perspect. 28 (3), 49–70.

Bartling, B., Fischbacher, U., 2012. Shifting the blame: on delegation and responsibility. Rev. Econ. Stud. 79 (1), 67–87.

Bartling, B., Fehr, E., Herz, H., 2014. The intrinsic value of decision rights. Econometrica 82 (6), 2005–2039.

Becker, G.M., DeGroot, M.H., Marschak, J., 1964. Measuring utility by a single-response sequential method. Behavioral Science 9 (3), 226–232.

Bénabou, R., Tirole, J., 2006. Incentives and prosocial behavior. Am. Econ. Rev. 96 (5), 1652–1678.

Bernheim, B.D., Taubinsky, D., 2018. Behavioral public economics. Handbook of behavioral economics: Applications and Foundations 1 (1), 381–516.

Bloom, N., Sadun, R., Van Reenen, J., 2012. The organization of firms across countries. The quarterly journal of economics 127 (4), 1663–1705.

Bobadilla-Suarez, S., Sunstein, C.R., Sharot, T., 2017. The intrinsic value of choice:The propensity to underdelegate in the face of potential gains and losses. J. Risk Uncertain. 54 (3), 187–202.

Boissonnet, N., Ghersengorin, A., 2022. Reactance: a freedom-based theory of choice.

Botti, S., Orfali, K., Iyengar, S.S., 2009. Tragic choices: autonomy and emotional responses to medical decisions. J. Consum. Res. 36 (3), 337–352.

Buffat, J., Praxmarer, M., Sutter, M., 2023. The intrinsic value of decision rights: a replication and an extension to team decision making. J. Econ. Behav. Organ. 209, 560–571.

Burger, J.M., Cooper, H.M., 1979. The desirability of control. Motiv. Emot. 3 (4), 381–393.

Cerreia-Vioglio, S., Dillenberger, D., Ortoleva, P., Riella, G., 2019. Deliberately stochastic. Am. Econ. Rev. 109 (7), 2425–2445.

Cettolin, E., Riedl, A., 2019. Revealed preferences under uncertainty: incomplete preferences and preferences for randomization. J. Econ. Theory 181, 547–585.

Chapman, J., Snowberg, E., Wang, S.W., Camerer, C., 2022. Looming Large or Seeming Small? Attitudes Towards Losses in a Representative Sample. National Bureau of Economic Research.

Chapman, J., Snowberg, E., Wang, S.W., Camerer, C., 2024. Dynamically Optimized Sequential Experimentation (DOSE) for Estimating Economic Preference Parameters. National Bureau of Economic Research w33013..

Chapman, J., Ortoleva, P., Snowberg, E., Yariv, L., Camerer, C., 2025. Reassessing Qualitative Self-Assessments and Experimental Validation. National Bureau of Economic Research.

Dean, M., Ortoleva, P., 2019. The empirical relationship between nonstandard economic behaviors. Proc. Natl. Acad. Sci. 116 (33), 16262–16267.

Deci, E.L., Ryan, R.M., 1985. The general causality orientations scale: self-determination in personality. J. Res. Personal. 19 (2), 109–134.

Deci, E.L., Ryan, R.M., 2000. The" what" and" why" of goal pursuits: human needs and the self-determination of behavior. Psychological Inquiry 11 (4), 227–268.

Deci, E.L., Ryan, R.M., 2006. Basic psychological needs scales.

Dekel, E., Lipman, B.L., Rustichini, A., 2001. Representing preferences with a unique subjective state space. Econometrica 69 (4), 891–934.

Dessein, W., Holden, R., 2022. Organizations with power-hungry agents. J. Law Econ. 65 (1), 263–291.

Dwenger, N., Kübler, D., Weizsäcker, G., 2018. Flipping a coin: evidence from university applications. J. Public Econ. 167, 240–250.

Fabrice, L.L., Tarroux, B., 2020. On attitudes to choice: some experimental evidence on choice aversion. J. Eur. Econ. Assoc. 18 (5), 2108–2134.

Fehr, E., Herz, H., Wilkening, T., 2013. The lure of authority: motivation and incentive effects of power. Am. Econ. Rev. 103 (4), 1325–1359.

Fehrler, S., Janas, M., 2021. Delegation to a group. Manag. Sci. 67 (6), 3714–3743.

Ferreira, J.V., Hanaki, N., Tarroux, B., 2020. On the roots of the intrinsic value of decision rights: experimental evidence. Games Econ. Behav. 119, 110–122.

Freundt, J., Herz, H., 2024. From partisanship to preference: how identity shapes dependence aversion. CESifo 11304.

Freundt, J., Herz, H., 2025. A cross-country study of preferences for choice autonomy. CESifo 12238.

Frey, B.S., Benz, M., Stutzer, A., 2004. Introducing procedural utility: not only what, but also how matters. J. Institutional Theor. Econ. 377–401.

Gillen, B., Snowberg, E., Yariv, L., 2019. Experimenting with measurement error: techniques with applications to the caltech cohort study. J. Polit. Econ. 127 (4), 1826–1863.

Gomez-Martinez, F., Herz, H., 2025. Who wants power? Measuring intrinsic preferences for power and their behavioral signature. CESifo 12241.

Hamilton, B.H., 2000. Does entrepreneurship pay? An empirical analysis of the returns to self-employment. J. Polit. Econ. 108 (3), 604–631.

Hofstede, G., 2001. Culture's Consequences: Comparing Values, Behaviors, Institutions and Organizations Across Nations. Sage.

Holt, C.A., Laury, S.K., 2002. Risk aversion and incentive effects. Am. Econ. Rev. 92 (5), 1644–1655.

Hurst, E., Pugsley, B.W., 2011. What Do Small Businesses Do? National Bureau of Economic Research.

Inglehart, R., et al., 2014. World Values Survey: Round Six-Country-Pooled Datafile Version. JD Systems Institute, Madrid. www.worldvaluessurvey.org/WVSDocumentationWV6.jsp.

Iyengar, S.S., Kamenica, E., 2010. Choice proliferation, simplicity seeking, and asset allocation. J. Public Econ. 94 (7–8), 530–539.

Kreps, D.M., 1979. A representation theorem for preference for flexibility. Econometrica 47 (3), 565–578.

Machina, M.J., 1985. Stochastic choice functions generated from deterministic preferences over lotteries. Econ. J. 95 (379), 575–594.

May, T., 2017. PM's Florence speech: a new era of cooperation and partnership between the UK and the EU.

Meemann, C., 2023. On the Economic Value of Decision Rights: An Experimental Test (Ph.D. dissertation). Universität der Bundeswehr Hamburg.

Mill, J.S., 1859. On Liberty. John W. Parker and Son, London, West Strand: 9.

Neri, C., Rommeswinkel, H., 2016. Freedom, power and interference.

Nussbaum, M.C., 2000. Women and Human Development: The Capabilities Approach, vol. 3. Cambridge University Press.

Owens, D., Grossman, Z., Fackler, R., 2014. The control premium: a preference for payoff autonomy. Am. Econ. J. Microecon. 6 (4), 138–161.

Pikulina, E., Tergiman, C., 2020a. Preferences for non-interference in pay. Working Paper..

Pikulina, E.S., Tergiman, C., 2020b. Preferences for power. J. Public Econ. 185, 104173.

Rotter, J.B., 1966. Generalized expectancies for internal versus external control of reinforcement. Psychological monographs: General and applied 80 (1), 1.

Scheibehenne, B., Greifeneder, R., Todd, P.M., 2010. Can there ever be too many options? A meta-analytic review of choice overload. J. Consum. Res. 37 (3), 409–425.

Schwarzer, R., Jerusalem, M., et al., 1995. Generalized self-efficacy scale. Measures in health psychology: a user's portfolio. Causal and control beliefs 1 (1), 35–37.

Sen, A., 1985. Commodities and Capabilities. Oxford University Press, Oxford.

Sethi-Iyengar, S., Huberman, G., Jiang, W., et al., 2004. How much choice is too much? Contributions to 401 (k) retirement plans. Pension design and structure: New lessons from behavioral finance 83, 84–87.

Sloof, R., von Siemens, F.A., 2017. Illusion of control and the pursuit of authority. Experimental Economics 20, 556–573.

Snibbe, A.C., Markus, H.R., 2005. You can't always get what you want: educational attainment, agency, and choice. J. Personal. Soc. Psychol. 88 (4), 703.

Stern, S., 2004. Do scientists pay to be scientists? Manag. Sci. 50 (6), 835–853.

Sugden, R., 2008. Capability, happiness and opportunity. In: Maurizio, P., Luigino, B., Comim, F., Chapter 13, (Eds.), Capabilities and Happiness, vol. 299. Oxford University Press.

Verme, P., 2009. Happiness, freedom and control. J. Econ. Behav. Organ. 71 (2), 146–161.

Wang, S.W., Filiba, M., Camerer, C.F., 2010. Dynamically optimized sequential experimentation (dose) for estimating economic preference parameters Unpublished Manuscript.