"The risks cannot be compensated": The willingness to donate DNA for science and its relationship with economic preferences

Karlsson Linnér, R.; Jain, M.

“The risks cannot be compensated”: The willingness to donate DNA for science and its relationship with economic preferences

Richard Karlsson Linnér
Department of Economics, Universiteit Leiden

Manisha Jain
Department of Economics, University of Wisconsin-Madison

Abstract. The accumulation of large genetic data is crucial for the scientific advancement of genetic research and precision medicine, but various participation biases threaten the validity of genetic research data sets. To better understand the decision to participate and its relationship with economic incentives and preferences, we studied the stated willingness to donate DNA for science by saliva sample in a representative panel of Dutch households. There were two randomized treatments, varying (i) the information material on benefits and risks and (ii) the intended financial incentive. The first treatment had no detectable effect, suggesting insensitivity to the information material. The higher incentive conditions had modest and diminishing effects, suggesting that offering higher incentives is not cost-effective. Stated reasons not to donate DNA concentrated on personal risks, e.g., privacy violations and data exploitation. Accordingly, stated risk willingness was found strongly associated, followed by trust and positive reciprocity. Revealed economic preferences were not associated. The results support previous findings for self-rated health, interpersonal trust and confidence in science or societal institutions but not for certain demographic variables (e.g., age, education and religiosity). We conclude by proposing strategies to encourage participation, e.g., to reallocate resources to risk-minimizing or compensatory measures.

Résumé. «On ne peut pas compenser les risques»: la volonté de donner de l’ADN à des fins scientifiques et son lien avec les préférences économiques. L’accumulation de grandes quantités de données génétiques est essentielle pour le progrès scientifique de la recherche génétique et de la médecine de précision, mais divers biais de participation menacent la validité des ensembles de données de recherche génétique. Afin de mieux comprendre la décision de participer et sa relation avec les incitations et les préférences économiques, nous avons étudié la volonté déclarée de donner de l’ADN à des fins scientifiques sous la forme d’un échantillon de salive chez un groupe représentatif de ménages néerlandais. Il y a eu deux traitements randomisés, variant (i) la documentation d’information sur les avantages et les risques et (ii) l’incitation financière prévue. Le premier traitement n’a pas eu d’effet détectable, ce qui suggère une insensibilité à la documentation d’information. On constate qu’il n’est pas rentable d’offrir des incitations élevées en raison de leurs effets modestes et décroissants. Les principales raisons invoquées pour ne pas faire de don d’ADN sont les risques personnels, comme les violations de la vie privée et l’exploitation

Corresponding author: Richard Karlsson Linnér, r.karlsson.linner@law.leidenuniv.nl

We are grateful to the study participants of the Dutch LISS Panel and the German Socio-Economic Panel for making this research possible. We thank Titus Galama, Hans van Kippersluis and Marcel Das for conducting survey no. 261 in the LISS panel and Eduardo Azevedo, Jonathan Beauchamp, Renato Frey and two anonymous reviewers for helpful feedback. This research was supported by the Dutch Research Council (NWO; project number VI.Veni.221E.080). The authors declare no conflicts of interest and are solely responsible for the views and opinions expressed in the manuscript. Section A1 reports extended acknowledgments and section A2 author contributions.

Canadian Journal of Economics / *Revue canadienne d’économique* 2025 58(2)
May 2025. / *Mai 2025.*

25 / pp. 515–547 / DOI: 10.1111/caje.70008

des données. Par conséquent, on constate un lien fort entre le don d'ADN et la volonté déclarée de prendre des risques, suivie de la confiance et de la réciprocité positive. On n'a pas établi de lien avec les préférences économiques exprimées. Les résultats confirment les constatations précédentes concernant la santé autoévaluée, la confiance interpersonnelle et la foi en la science ou envers les institutions sociétales, mais pas pour certaines variables démographiques (comme l'âge, l'éducation et la religiosité). Nous concluons en proposant des stratégies pour encourager la participation, par exemple en réaffectant des ressources à des mesures de réduction des risques ou à des mesures compensatoires.

JEL classification: C90, D61, I18

1. Introduction

"The risks cannot be compensated and the compensation is small."

–A panel member on what prevents them from donating DNA for science

Recent advances in genetic research were accelerated by technological innovations and massive data generation (Zeggini et al., 2019). For example, the UK Biobank created a major scientific catalyst by collecting genetic material (DNA) from half a million volunteers (Glynn and Greenland, 2020). Regardless, tens of millions more samples are needed to realize the potential of precision medicine (O'Connor, 2021). Governments are investing billions (in US dollars) in collecting more genetic data (The All of Us Program, 2019), which emphasizes the policy relevance of understanding the decision to participate. It is crucial to improve the participation rate, particularly for underrepresented disease cases and social groups (Wang et al., 2023). This study investigates the willingness to donate DNA for science in the Dutch LISS panel, focusing on its relationship with financial incentives, economic preferences and socioeconomic characteristics.

The increasing importance of genetic data in economic research, e.g., to investigate gene–environment interactions or to create instrumental variables, further motivates studying the willingness to donate DNA as a potential source of bias (Fletcher, 2018; Biroli et al., 2022; Benjamin et al., 2024). Representative sampling is important for study validity and unbiased estimation; but volunteer, non-response and other participation biases are documented problems of existing genetic data sets (Schoeler et al., 2023). In the example of the UK Biobank, participants tend to be healthier, wealthier, more educated and female (Fry et al., 2017). Sampling bias reduces the statistical power, increases the false positive rate and can bias analyses with instrumental variables (Munafò et al., 2018; Hughes et al., 2019). A better understanding of the decision to participate may help improve the recruitment process or prevent systematic sample attrition (Walter and David, 2016). This knowledge could also guide stratified sampling or post-stratification procedures (Hughes et al., 2019), e.g., by highlighting variables to use in variations of the renowned Heckman (1979) selection model, which was recently extended to genome-wide data analysis by Pirastu et al. (2021).

There is some previous research on people's attitudes toward donating DNA for science, most notably a series of papers based on the Global Alliance for Genomics and Health (GA4GH) survey of an impressive 36,000 people in 22 countries, called "Your DNA, Your Say" (Middleton et al., 2018, 2020). This effort surveyed the stated willingness to donate DNA and medical information anonymously to a database (i.e., data donations only), conditional on key user types (medical doctors, non-profit researchers or for-profit researchers). Another focus was to survey familiarity with genetics and opinions about genetic exceptionalism (i.e., whether genetic data merits a special status distinct from other personal health information). The survey collected a handful of demographic variables, such as age, sex, family structure, education, religiosity and trust (Middleton et al., 2019; Milne et al., 2019; Voigt et al., 2020).

The authors report that the overall willingness to donate DNA and medical data was modest—only about 50% (Middleton et al., 2020). The willingness was similar across countries, but significantly greater (67%) in the English-speaking group (Middleton et al., 2019). People were noticeably more skeptical of for-profit researchers, while the willingness was similar between medical doctors and non-profit researchers (Middleton et al., 2020). Familiarity with genetics and opinions of genetic exceptionalism mostly predicted greater willingness, as did some demographic variables (Middleton et al., 2019; Voigt et al., 2020). Below, we review their findings in more detail and compare them with ours. The present study contributes by studying a similar decision process in a country that was not surveyed—the Netherlands—and by exploring additional plausible explanatory factors of interest to the social sciences and policy research, such as incentives and preferences.

1.1. The present study

Here, we study the stated willingness to donate DNA for science by saliva sample in the LISS panel, a representative panel of Dutch households that surveys a rich collection of demographic and social-scientific variables, including experimentally elicited preferences (Scherpenzeel and Das, 2010). The economic preferences under study are listed in table 1. Most are stated self-assessment items that correspond well with the Global Preference Survey (Falk et al., 2018), but some are revealed measures elicited with real monetary rewards (on top of the standard remuneration of €15 per survey hour). The research strategy was preregistered, specifying the aims and research design, variable definitions, power calculations, multiple-testing correction, etc.[1]

In November 2020, the LISS panel surveyed the stated willingness to donate DNA for science by saliva sample. The ultimate purpose of this prospective collection of a biospecimen and extraction of DNA data was described as non-commercial social, scientific and public health research. Respondents were randomized into two experimental treatments. The first treatment varied the amount of information shown in the preamble on potential benefits and risks (∼100 words extra in addition to a general 350-word introduction). The second treatment varied the hypothetical financial incentive (showing €10, €20 or €50). Thus, survey no. 261 measures the *stated* willingness to donate DNA to a potential future data collection for non-commercial research, but no genetic samples have yet been collected there. Additionally, we report a replication analysis of *revealed* donations of DNA in the German Socio-Economic Panel (SOEP), which were collected without extra incentives (Koellinger et al., 2023). Section 2 reports further details on the data sources and research strategy.

The main empirical analysis in section 3 consists of a series of logistic regressions, step-wise evaluating sets of explanatory variables, while retaining a satisfactory sample size. The only experimental treatment with a detectable effect was the €50 condition. However, despite strong significance, its average marginal effect was modest (AME = 6.1 percentage points, pp). The €20 condition had slightly larger effect in proportion to the amount (3.3 pp), but it was not significant. These results suggest diminishing returns to higher incentives and limited encouraging effects. In section 3.2.2, we return to this question and show that higher incentives have limited cost-effectiveness. Because the experiment was only hypothetical, our results establish, at least, a plausible lower bound on the effect of real monetary rewards.

1 The preregistered analysis plan is available at `https://osf.io/8tzp9`.

TABLE 1
Summary of stated and revealed economic preferences

Stated measures	Brief item description	Scale	N
Patience	Two repeated measures of the willingness to forego something today for the possibility of a future benefit	0–10	4,728
Risk willingness	Nine repeated measures of the general willingness to take risks	0–10	5,022
Positive reciprocity	Sum score over three questions (no repeated measures) on the willingness to reciprocate favours and help from others	0–12	3,078
Negative reciprocity	Sum score over three questions (no repeated measures) on the willingness to reciprocate unjust treatment or insults from others	0–12	3,078
Altruism (donation)	Three repeated measures of having donated to humanitarian, environmental, peace or animal rights organizations in the last 12 months	0–1	5,304
Altruism (helpful)	Three repeated measures of rating being helpful to be an important guiding life principle	1–7	5,245
Trust	Three repeated measures of whether most people can be trusted versus you can't be too careful in dealing with people	0–10	5,246
Revealed measures	**Brief item description**	**Scale**	**N**
Risk willingness	Percentile rank in a staircase procedure of binary lotteries where subjects decide between a risky and safe bet (average of up to four experiments)	0–100	2,374
Ambiguity aversion	Rank in a staircase procedure where subjects decide between a lottery with known probabilities versus a lottery with unknown probabilities (single experiment)	0–62	772
Altruism (donation)	Whether decided to donate part or all of a lottery prize to charity (single experiment)	0–1	1,360
Trust	Percentile rank of the belief that other players are pro-social (single experiment)	0–100	557

NOTES: The table design was inspired by Falk et al. (2018). Sample means and standard deviations are reported in table 3. Additional details are reported in sections A7 and A8 and in tables S2 to S4.

Reassuringly, many studies report convergence between fictive and real choice experiments (Campos-Mercade et al., 2021; Brañas-Garza et al., 2023).

Next, among the economic preferences, stated risk willingness was found to be the strongest and most consistent explanatory factor, predicting 4.4 pp greater willingness per standard deviation (SD) unit. This result reverberated with the participants' open-ended text responses (analyzed in section 3.3) that concentrated on personal risks, such as privacy violations, data breaches and exploitation by third parties (including insurance companies, the police, approved researchers and the government). The association with stated risk willingness was supported by the replication analysis.

Moreover, stated measures of trust and positive reciprocity were found suggestively associated, with marginal effects about half as large as risk willingness, but they did not replicate at the 5% level. In line with previous research (Frey et al., 2017; Hertwig et al., 2019), revealed economic preferences converged poorly with stated counterparts and did not explain a significant amount of variation. Lastly, we found no appreciable evidence of association with measures of altruism, nor that higher amounts crowded out such motivates. These findings align with Abeler and Nosenzo (2015) and Thiemann et al. (2022) that found no effect of appealing to altruistic motives on participation in economic experiments.

Our main recommendation to encourage participation, among those we propose in section 5, is to shift resources away from small incentives to instead prioritize risk-minimizing or compensatory measures that may better offset the perceived risks. Our main association with stated risk willingness, and the panel members' emphasis on privacy violations in their text responses, both support the notion that many people think this decision exposes them to various personal risks. A remedy could be to prioritize available funds for a liability insurance or trust fund, that could pay out large compensatory damages. A direction for future research is to investigate whether insurance can offset the perceived risks better than small financial incentives.

1.2. Research participation in the economic literature

Our results contribute to the economic literature on the relationship between preferences and participation in research or lab experiments. Roe et al. (2009) show among university students that the stated willingness to donate DNA for science (by blood sample) was negatively associated with stated risk aversion, but not with its revealed counterpart, which is essentially what we found. On the contrary, once adjusted for covariates, Slonim et al. (2013) found no effect of neither stated nor revealed risk preferences on the decision to attend a conventional but lengthy economic experiment. They did, however, find associations with altruism. In other student populations, Cleave et al. (2013) and Thiemann et al. (2022) report no differences in the participation rate by pro-social or risk preferences. Thus, the evidence on the impact of economic preferences on research participation in student populations is mixed. Our research set-up and findings are most similar to Roe et al. (2009) and our results extend the evidence on risk preferences to the general population.

Further, Harrison et al. (2009) show in a large Danish population sample that lottery-based incentives (rather than fixed amounts) can induce self-selection on risk preferences. Their findings, combined with ours, suggest that lottery-based incentives are inadvisable in genetic research because they could intensify the selection we found on risk willingness. Lastly, while our results suggest that higher incentives do matter somewhat, we do not find such a drastic effect as do Abeler and Nosenzo (2015). They report a two thirds reduction in the participation rate when no reward was mentioned in the invitation to a lab experiment. Our replication data also contradict a drastic reduction because the overall willingness was found to be comparable without a financial incentive. Beyond the economic literature, our results contribute to the growing literatures on public opinions about donating DNA and on sampling biases in genetic data sets (Hughes et al., 2019; van Alten et al., 2023; Schoeler et al., 2023).

1.3. Paper structure

The remainder of the article is structured as follows. Section 2 describes the data sources, survey no. 261, variable definitions and our research strategy. Section 3 reports the main regression analysis to investigate the role of incentives, economic preferences and other socioeconomic characteristics. The regression results are then used in section 3.2.2 to analyze the cost-effectiveness of the financial incentives. Section 3.3 analyses the panel members' open-ended text responses. In section 4, we extend a theoretical model to stylize the decision to donate DNA as a risky decision. Section 5 discusses the relevance of our findings and proposes strategies that can encourage participation. Appendix sections A3 to A13 report additional details on the data and study procedure.

2. Data and research strategy

This study sourced its primary data set from the LISS panel (Longitudinal Internet Studies for the Social Sciences), administered by Centerdata (Tilburg University) (Scherpenzeel and Das, 2010). The LISS is a household panel study based on a representative sampling frame provided by Statistics Netherlands (Scherpenzeel, 2011). The data are typically described as representative (Drerup et al., 2023), which we verified (appendix A5). To reduce volunteer effects, membership is by invitation only. Since its inception in 2007, the panel tracks about 5,000 households with replacement for attrition (∼7,000 active respondents per annual wave). The informed consent restricts data access to non-commercial scientific or policy-relevant research only.[2]

Each month, a designated household contact person gets asked to confirm the accuracy of (or record changes to) a standard set of demographic and household composition measures, which must be completed before the household members can participate in any other surveys that month. The annual core survey modules are Health; Religion and Ethnicity; Social Integration and Leisure; Family and Household; Work and Schooling; and Personality, Politics and Values; as well as the three core surveys Economic Situation: Assets (biennial); Economic Situation: Income; and Economic Situation: Housing. In addition to the core surveys, hundreds of one-time surveys and experimental modules have been assayed over the years. Below, we describe how we selected variables from these sources, most of which were sourced from the core modules (except the economic preference variables, which were mostly sourced from various experimental modules).

2.1. Details on the survey "Willingness Genotyping"

Our analyses take survey no. 261 ("Willingness Genotyping") as the main sampling frame, to which all 6,832 active members in October 2020 (age 16+) were invited.[3] The exact vignette, questionnaire and variable codebook are publicly available in the LISS data archive.[4] In total, 5,366 respondents (78.5% of 6,832) completed this questionnaire in November 2020 (five were later dropped because they were missing from the background data). To our knowledge, this is the first study on survey no. 261 and possibly the first to observe and correct for non-response bias in the context of donating DNA.

The survey was short and straightforward. All participants were first presented a vignette with general background information (350 words) on the goals of the genetic research (i.e., to promote research on the health and financial situation of the Dutch population) and the possible benefits and risks of donating a saliva sample and DNA data. Half of the sample was randomized in the first experimental treatment to be presented with more information on benefits and risks (100 words extra). The second experimental treatment randomized the hypothetical financial incentive shown at the end of the preamble (€10, €20 or €50). We examined for sample imbalance and confirmed successful randomization (table S1).[5]

2 The LISS data can be accessed (free of charge) at https://www.lissdata.nl/.

3 Our sample of non-responders includes two extra members that were eligible but eventually not invited to survey no. 261 (perhaps due to intermittent withdrawal). The data could not distinguish these for exclusion. Their inclusion should have a negligible impact on the non-response weighting.

4 English and Dutch versions of the documents can be accessed at https://dataarchive.lissdata.nl/study-units/view/1237.

5 Supplemental tables (S1 to S13) are available at https://osf.io/32dnr.

Our interpretation of the general background information is that it was focused on communicating the societal benefits (e.g., improving public health), the non-invasive saliva sampling procedure, the de-identification procedure and explaining that the security measures would follow European regulatory standards. The information material did not mention the scheduled destruction of any donated materials, so we assume the respondents considered donating both a biospecimen and DNA data to be kept long term under broad consent. The informed consent information material specifies that the data can be accessed only for non-commercial research. In this way, survey no. 261 departs from "Your DNA, Your Say," which surveyed donating DNA and medical data (but no biospecimen) conditional on key user types.

We interpret the extra information in the first experimental treatment to be focused on privacy risks to the individual, while benefits were presented in terms of their societal importance. Importantly, the general information stated clearly that the respondents themselves would not benefit directly from participating in the research. Because additional benefits *and* risks were presented in tandem, we did not have a clear expectation about the direction of effect but hypothesized that more information should increase the willingness by reducing uncertainty. Potentially, the net effect on the willingness could also cancel out or be truly zero, which we could not test with these data. Based on expected utility theory, the larger incentive amounts in the second experimental treatment were hypothesized to increase the willingness.

The survey preamble concluded with displaying the randomized incentive amount, which was followed by a single question on the likelihood of participating, measured on a five-point scale:

(1) Certainly participate
(2) Probably participate
(3) Perhaps participate or perhaps not participate (50/50 likelihood)
(4) Probably not participate
(5) Certainly not participate

Participants that did not answer "certainly participate" were prompted with a follow-up question asking for an open-ended text response ("in a number of words or sentences") about what holds them back.

2.2. Variable definitions

The preregistration specified that the stated likelihood of participating would be dichotomized so that answers 1–2 were coded as "willing to donate DNA for science by saliva sample" (1; $N = 2{,}857$) and answers 3–5 as "unwilling" (0; $N = 2{,}509$), which resulted in a roughly even split (53.2% willingness). We later found support for a roughly even split in the replication data. Specifically, the revealed willingness in the SOEP turned out to be similar (57.1%, table S1b). Below we analyze the dichotomized willingness to donate DNA with logistic regression. In robustness checks, we complemented the analysis with ordinal logistic regressions of the five-point coding (reversed to measure greater willingness), which showed no meaningful differences.

2.2.1. Variable selection and sources

This study is cross-sectional because survey no. 261 has been assayed only once (and no genetic data have been collected since). Appendix A3 describes how we searched the LISS codebooks to identify variables for the preregistration, which was done without dipping into the data. Eventually, we had preregistered a curated selection of about 75 candidate variables. The preregistration did not specify causal hypotheses for the candidate variables because they were not exogenous (see section A4).

Appendix A6 describes how we leveraged the longitudinal data to impute missing values and to reduce measurement error by averaging certain repeated measures. Unfortunately, we later found it infeasible to fit most candidate variables in the same regressions model because of non-overlapping missingness. Instead, we proceeded by defining a baseline model on which we evaluated variable sets that were conceptually meaningful to fit together while retaining a satisfactory sample size.

Ultimately, the rich survey data enabled us to code seven measures of stated economic preferences and four measures of revealed economic preferences (table 1). With the exception of stated trust and stated altruism, which were sourced from a core survey (no. 7, Personality), the economic preferences were all sourced from different experimental modules (listed in tables S2–S3). Appendix sections A7 to A8 report further details on the economic preferences.

Most other explanatory variables were sourced from the core modules and have straightforward definitions (reported in table S1a). In terms of timing, the core modules were surveyed over the year 2020 and all were assayed prior to survey no. 261. The revealed preferences were also all assayed beforehand, with the exception of revealed altruism (donation) assayed in a 2022 survey. The variables on stated risk willingness, trust and altruism were available for almost the entire sample, while the other economic preferences were observed only in subsets (see table 1).

2.3. Descriptive statistics and non-response weighting

Sample descriptive statistics are reported separately for responders and non-responders to survey no. 261 in tables 2 to 3. The following baseline model of control variables was defined using standard demographic variables with little missingness (5,037 complete observations): age, sex, ever married, urban residence, education, paid work, on unemployment benefits (UB) or income support, religiosity, raised within a certain faith and two dummies for non-Western and/or immigrant background. We first considered income (either personal or household), but this variable was eventually excluded because of additional missingness and no evidence of association.

We evaluated the possibility of systematic non-response. The response rate was noticeably lower by younger age, paid work and immigrant and/or non-Western background. Importantly, however, there were no marked differences by any of the economic preferences (see table 3). Nevertheless, all regressions below were adjusted by inverse-probability weighting. To generate the non-response weights, we ran a probit regression of accepting the invitation to survey no. 261 on the baseline covariates ($N = 6{,}148$ of which 5,037 had accepted the invitation). The variables age, sex, paid work and extent of religiosity were significant, while primary school education and immigrant and/or non-Western background were not, though the signs and magnitudes of their coefficients aligned with their raw proportions. The McFadden pseudo R^2 was 6.3%, suggesting the model explained a meaningful amount of the variation in the non-response (table S6). Lastly, we checked that no respondent was assigned an extreme weight.

2.4. Replication data

The German SOEP recently enriched their representative Innovation Sample (IS) panel by collecting DNA. The data collection effort is described by Koellinger et al. (2023) and its purpose and execution corresponds well with the prospective survey in LISS. These data, which

TABLE 2
Descriptive statistics: Baseline covariates

	Responders		Non-responders	
Variable	Mean (SD), %	*N*	Mean (SD), %	*N*
Age	52.6 (18.6)	5,361	41.1 (16.6)	1,473
Sex (1 = Female)	53.3%	5,361	57.2%	1,473
Ever married	67.7%	5,361	49.1%	1,473
Urban residence	39.3%	5,340	44.1%	1,461
Non-Western background	9.1%	5,220	15.7%	1,397
Immigrant background	18.9%	5,220	26.6%	1,397
Paid work	48.4%	5,361	63.3%	1,473
Income support	3.9%	5,361	3.2%	1,473
Diploma	**100.0%**	**5,361**	**100.0%**	**1,473**
1. Primary school	7.4%	397	10%	148
2–3. High school	29.4%	1,578	24%	353
4. Intermediate vocational	23.8%	1,278	23.7%	349
5. Higher vocational	26.1%	1,398	26.6%	392
6. University	13.2%	710	15.7%	231
Religious	**100.0%**	**5,219**	**100.0%**	**1,180**
1. Certainly	11.7%	611	13.5%	159
2. Somewhat	20.4%	1,065	18.7%	221
3. Barely	26.5%	1,382	26.5%	313
4. Certainly not	41.4%	2,161	41.3%	487
Raised faith	**100.0%**	**5,257**	**100.0%**	**1,185**
None	37.7%	1,980	44.6%	528
Roman Catholic	32.1%	1,687	24.7%	293
Other Christian	26.1%	1,374	24.1%	285
Non-Christian	4.1%	216	6.7%	79
Net monthly income (euro)	1,847.0 (3,654.3)	5,088	1,665.0 (1,146.4)	1,379

NOTES: Table S1 reports the complete sample descriptive statistics. The Dutch educational levels VMBO (US: junior high school) and HAVO/WVO (US: senior high school) were merged into a single category (2–3. High school). Net monthly income was later dropped from the baseline model because of more than 5% missingness and no evidence of association. Bold = 100%.

were publicized after the preregistration, thus provided a newfound opportunity for replication, but this time with revealed donations. The SOEP is representative household panel of a neighbouring European country of reasonably similar culture and economic development (Richter and Schupp, 2015). These data were used in a replication analysis of the models with stated economic preferences (table 4). Pursuing other analyses with the replication data was deemed outside our scope.

Briefly, all SOEP-IS households that participated in the 2019 wave were invited to donate DNA for science by buccal swab in an interview.[6] A total of 4,182 panel members (17+ years) were presented with the background information, followed by the invitation and consent procedure. Notably, the invitation was given in face-to-face interviews and the information material included an encouraging video. Eventually, 2,394 adults provided consent and saliva samples (57.1%, table S1b). Importantly, beyond the standard remuneration, there was no extra financial incentive, which provides us the opportunity to analyze *revealed* donations

6 Saliva samples and buccal swabs are comparable non-invasive procedures to donate DNA. We expect no meaningful impact on the study results from this distinction.

TABLE 3
Descriptive statistics: Selected variables of interest

	Responders		Non-responders	
Variable	Mean (SD), %	*N*	Mean (SD), %	*N*
Donate DNA for science	**100.0%**	**5,361**	Missing for non-responders to survey no. 261	
1. Certainly not	15.6%	835		
2. Probably not	15.0%	804		
3. Perhaps (50/50)	16.2%	870		
4. Probably	25.4%	1,361		
5. Certainly	27.8%	1,491		
Dichotomized (4–5 = "Willing")	53.2%	5,361		
Stated preferences				
Patience (0–10)	6.7 (1.9)	4,728	6.6 (2.0)	743
Risk willingness (0–10)	4.8 (1.8)	5,022	4.8 (1.9)	1,002
Positive reciprocity (0–12)	9.4 (2.2)	3,078	9.4 (2.1)	514
Negative reciprocity (0–12)	4.8 (2.8)	3,078	4.6 (2.8)	514
Altruism (donation)	28.4%	5,304	26.9%	1,048
Altruism (helpful) (1–7)	5.9 (0.8)	5,245	5.9 (1.0)	1,118
Trust (0–10)	6.1 (2.0)	5,246	6.0 (2.1)	1,133
Revealed preferences				
Risk willingness (0–100)	27.6 (25.9)	2,374	26.6 (25.6)	369
Ambiguity aversion (0–62)	26.5 (15.0)	772	25.1 (12.8)	99
Altruism (donation)	66.0%	1,360	66.0%	144
Trust (0–100)	22.1 (22.2)	557	18.1 (16.9)	84
Other domains				
Agreeableness (0–40)	28.5 (4.9)	5,248	28.3 (5.2)	1,123
Conscientiousness (0–40)	27.3 (4.9)	5,248	26.0 (5.3)	1,123
Extraversion (0–40)	22.0 (6.5)	5,248	22.4 (6.5)	1,123
Neuroticism (0–40)	15.2 (6.9)	5,248	16.6 (7.0)	1,123
Openness (0–40)	25.0 (4.8)	5,248	25.6 (4.8)	1,123
Self-esteem (0–60)	45.1 (9.8)	5,247	43.3 (10.7)	1,120
Overall health	**100.0%**	**5,318**	**100.0%**	**981**
1. Poor or moderate	17.2%	917	16.5%	162
2. Good	55.2%	2,935	54.3%	533
3. Very good or excellent	27.6%	1,466	29.2%	286
Government may store DNA?	**100.0%**	**1,957**	**100.0%**	**303**
1. Absolutely never	15.6%	306	15.8%	48
2. Strict circumstances	58.9%	1,153	62.0%	188
3. Certainly permissible	25.4%	498	22.1%	67

NOTES: Table S1 reports the complete sample descriptive statistics. Bold = 100%.

of DNA for science in the absence of direct incentives. Appendix A13 reports details on the replication data.

3. Analysis of the willingness to donate DNA for science

3.1. Descriptive evidence

The stated willingness to donate DNA for science was 53.2% in the LISS (table 3). This proportion was not statistically different from a previous estimate of 52.9% stated willingness (to donate DNA and medical data to non-profit research) that was surveyed in the neighbouring country Belgium (Middleton et al., 2020). Similarly, the *revealed* willingness that we report for the German SOEP (57.1%, table S1b) was not significantly different from a previous estimate of 56.0% *stated* willingness (to donate DNA and medical data to any user type) in Germany (Voigt et al., 2020).

TABLE 4
Selected results: Logistic regression analysis

Variable	Model 1	Model 2a	Model 2b	Model 2c
Age	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)
Sex (1 = female)	−0.07 (0.05)	−0.09 (0.06)	−0.03 (0.06)	0.06 (0.08)
Ever married	0.01 (0.08)	0.04 (0.09)	0.03 (0.09)	0.06 (0.12)
Urban residence	0.12 (0.07)	0.14 (0.07)*	0.12 (0.07)	−0.01 (0.09)
Non-Western background	−0.17 (0.16)	−0.16 (0.16)	−0.21 (0.17)	−0.25 (0.25)
Immigrant background	0.08 (0.10)	0.07 (0.10)	0.10 (0.10)	0.11 (0.14)
Paid work	−0.06 (0.07)	−0.07 (0.07)	−0.05 (0.07)	−0.07 (0.10)
Income support	0.22 (0.15)	0.24 (0.15)	0.25 (0.16)	0.22 (0.20)
Diploma (vs. high school)				
Primary school	−0.16 (0.12)	−0.17 (0.13)	−0.23 (0.13)	−0.38 (0.19)*
Interm. vocat.	0.08 (0.08)	0.07 (0.08)	0.06 (0.09)	0.10 (0.11)
Higher vocat.	0.05 (0.08)	0.03 (0.08)	0.00 (0.09)	−0.11 (0.11)
University	−0.37 (0.10)***	−0.40 (0.11)***	−0.42 (0.11)***	−0.46 (0.14)**
Religious (vs. Certainly)				
Somewhat	−0.01 (0.11)	0.01 (0.11)	0.03 (0.12)	0.17 (0.15)
Barely	0.21 (0.11)	0.20 (0.11)	0.21 (0.12)	0.29 (0.15)*
Certainly not	0.30 (0.11)**	0.33 (0.11)**	0.34 (0.12)**	0.41 (0.15)**
Treatments				
Extra information	0.06 (0.06)	0.06 (0.06)	0.07 (0.06)	0.07 (0.08)
Incentive (vs. 10 euro)				
20 euro	0.14 (0.07)	0.13 (0.07)	0.12 (0.07)	0.07 (0.09)
50 euro	0.25 (0.07)***	0.26 (0.07)***	0.28 (0.07)***	0.16 (0.09)
Stated preferences				
Altruism (donate)		0.02 (0.07)	0.00 (0.07)	−0.03 (0.10)
Altruism (helpful)		0.08 (0.04)*	0.07 (0.04)	−0.01 (0.05)
Trust		0.04 (0.02)*	0.04 (0.02)*	0.06 (0.02)**
Risk willingness			0.09 (0.02)***	0.10 (0.03)***
Patience				0.03 (0.02)
Positive reciproc.				0.05 (0.02)**
Negative reciproc.				0.00 (0.01)
Intercept	−0.22 (0.19)	−0.86 (0.30)**	−1.46 (0.33)***	−2.00 (0.48)***
Additional vars.	Yes	Yes	Yes	Yes
N	5,037	4,914	4,649	2,852
R^2	1.0%	1.2%	1.7%	2.2%

NOTES: Standard errors are reported in parentheses. “Additional vars.” indicates whether the regression model included additional variables omitted from the table. The complete results are reported in tables S7 to S12. Please note that model 2c in this table corresponds to table S8 panel D1 (and not C1). Significance was evaluated at the Bonferroni-corrected significance threshold $P \leq 0.001$ and suggestive significance at an FDR-corrected $P \leq 0.05$ (Benjamini−Hochberg). The stars $*$, $**$ and $***$ denote an (uncorrected) P less than 0.05, 0.01 and 0.001, respectively.

The descriptive evidence suggests, in line with previous research, that the overall willingness to donate DNA for science is modest. Our study extends this result to the panel data setting, where the participants could otherwise be more accepting or experienced with sharing sensitive information with researchers. Reassuringly, our comparison between the stated and revealed willingness in Germany suggests that hypothetical and real choice experiments appear to converge well for the decision to donate DNA for science.

Next, both experimental treatments were hypothesized to induce greater willingness. Turning to the first treatment, respondents randomized to the extra information condition were only slightly more willing, 53.95% vs. 52.4%, but this difference was not significant. The second treatment, which varied the incentive amount, also induced greater willingness:

56.2% (€50), 53.3% (€20), vs. 50.1% (€10). The difference between €50 and €10 was strongly significant, while the intermediate steps were not. Overall, these fairly small differences suggest a general insensitivity to both treatments and that we can expect to find limited treatment effects in the regression analysis, reported next.

3.2. Logistic regression analysis

This section reports our series of regression analyses of the stated willingness to donate DNA for science, where this outcome was analyzed as a binary variable (y) with a logistic regression model:

$$Pr(y = 1|\mathbf{X}) = \frac{\exp(\mathbf{X}\beta)}{1 + \exp(\mathbf{X}\beta)}, \tag{1}$$

where $\mathbf{X}$ is an $N \times K$ matrix holding an intercept column and $K - 1$ explanatory variables, with its corresponding $K \times 1$ parameter vector β. With this general framework, we estimated a series of regressions, varying the sets of regressors represented by $\mathbf{X}$. The regressions were weighted using the inverse-probability weights described above in section 2.3. Heteroskedasticity-consistent and cluster-robust standard errors were grouped at the household level to account for this dependency.

We had preregistered a stringent Bonferroni-corrected significance threshold of 0.001 (assuming 50 independent tests). Because many tests were expected to be correlated, as a complement, the preregistration specified a suggestive significance threshold at a Benjamini−Hochberg false-discovery rate (FDR) adjusted $P < 0.05$.

To check the robustness of our results to the preregistered decision to dichotomize the dependent variable, all model specifications were also estimated using ordered logistic regression. There were no meaningful differences in the estimates (tables S7−12), suggesting robustness. Also, below we report AMEs computed over the entire sample. As a robustness check, AMEs were computed for the bottom group (i.e., no extra information and €10 incentive), which resulted in basically identical estimates (tables S7−12). Therefore, the presentation below reports only the full-sample AMEs.

3.2.1. Results for the experimental treatments

The first regression fitted the baseline model together with the two experimental treatments (model 1, table 4). Most of the baseline variables were not significantly associated, with the exception of (i) religiosity (4. Certainly not religious, AME = 7.4 pp vs. 1. Certainly religious) and (ii) education (5. University education, AME = −9.2 pp vs. High school). We discuss these findings further in section 3.2.4.

The coefficient of the first treatment (i.e., extra information) was not statistically significant, which held across all model specifications and its effect was small (AME = 1.5 pp). We interpret this finding as that slight variation (∼100 words) in the information material did not have an impact large enough to be detected in a fairly well-powered sample. This finding is an interesting contrast to some studies on survey methodology that suggest strong response bias already from trivial differences in wording (Grover and Vriens, 2006). An alternative explanation of our null finding is that the effects of extra information on benefits vs. risks are opposing and cancelled out, but the data did not allow us to test this hypothesis.

Next, the €50 condition in the second treatment was strongly significant, while its effect was modest (AME = 6.1 pp) compared to other variables, such as education or religiosity. The €20 condition had slightly larger effect in proportion to the amount (AME = 3.3 pp), but this coefficient was not significant. Because we consider it reasonable to assume that real financial incentives should produce similar or larger effects, we consider our estimates to establish, at least, a plausible lower bound. Overall, the regression analysis suggests insensitivity to the higher incentives, though the highest amount effectively decreased the size of the unwilling group by 12.4%.

The experimental treatments were hypothesized in the preregistration to potentially interact with selected variables of interest (e.g., altruism, income or risk willingness). Therefore, a series of preregistered interaction terms were tested for association at the less stringent $P < 0.05$ (because of the reduced power to detect interactions), but none of the tested interaction terms was significant (results omitted). An ex post power analysis, which could now be informed by the largest effect estimated for the experimental treatments (i.e., €50), showed insufficient power (6.6% to 15.5%) to detect reasonably sized interactions (a 10% to 25% moderation of effect). To achieve 80% power, a 65% moderation of effect would be required. Overall, there was no evidence of any interactions with the experimental treatments and we did not test for interactions between any other explanatory variables.

3.2.2. The cost-effectiveness of the financial incentives

Deciding the right incentive amount to maximize the response rate, while minimizing the study costs, presents an interesting optimization problem for economic analysis. This situation shares some similarities with a monopsony employer, but instead of having to pay current workers a higher wage to attract more labour, the administration has to pay already-accepting respondents a larger-than-necessary reward to attract more participants.

To establish a baseline response rate under the €10 incentive, we computed the average fitted probability within the €10 group as predicted by model 1, which was $Pr_{€10} = 50.3\%$. The AMEs reported above imply that $Pr_{€20}$ and $Pr_{€50}$ are 53.6% and 56.4%, respectively. Let K denote the number of invitations sent and m the incentive amount. In this case, the expected number of accepted invitations conditional on m is $E[N] = K \times Pr_m$ and the expected cost of the study is $E[C] = N \times m$. It follows that the number of invitations required to expect a given target sample size is N/Pr_m.

Let us first evaluate these expressions in a situation where K has an upper bound, like the number of active panel members. In this case, a low acceptance rate cannot be compensated for by simply sending more invitations (for now, assumed to cost €0). If we set the upper bound to the size of the active panel (i.e., 6,832), then the expected number of responses for the three amounts are 3,436, 3,662 and 3,853, respectively, for the €10, €20 and €50 amounts. Their respective expected total costs are €34,360, €73,240 and €192,650. In other words, the substantial extra costs from offering the €50 incentive (€158,290) may attract only as few as 417 extra participants. The average cost per such marginal participant is, thus, about €380 (or €172 under the €20 incentive). Thus, attracting additional respondents by raising the incentives for everyone has a high price.

Next, if there is no clear upper bound on K (except population size), e.g., in the planning stages of a new biobank, then it is possible to compensate for the response rate by sending more invitations. Obviously, if the invitation has no cost, then it will always be cheaper to simply send more invitations. So, let us introduce a non-zero administration and material cost per invitation (P). Then, for a given target N, a break-even point between a low (L) and high (H) incentive can be defined as

$$(K_L \times P) + (N \times m_L) = (K_H \times P) + (N \times m_H). \tag{2}$$

To expect a target sample size of, say, 10,000 accepted invitations, the above expectations suggest we must send 19,881 (€10 incentive), 18,657 (€20 incentive) and 17,730 (€50 incentive) invitations. Next, we solve the break-even equation for P:

$$P = N \times (m_H - m_L)/(K_L - K_H), \tag{3}$$

followed by plugging in our estimates. We find that an administration and material cost of $P \geq$ €82 is required for the €20 incentive to break even with €10, while the €50 incentive does not break even until $P \geq$ €186.

In conclusion, the above shows that the costs from offering higher incentives require either a limited sampling frame or a considerable administration and material cost to be motivated. Generally, the smaller amounts seem the most cost-effective and our replication analysis suggests little harm from no financial incentive. The modest effects we estimated for the incentive conditions align with a systematic review of real clinical trials (Huynh et al., 2014), which found only a small effect of *real* monetary incentives. Our results support to their conclusion that it is generally more cost-effective to send more invitations (or to enhance their quality).

3.2.3. Results for the Economic Preference domain

The results for the stated economic preferences are reported in table 4. The first model fitted trust and two measures of altruism on top of the baseline model (N dropped from 5,037 to 4,914). Next, risk willingness was added ($N = 4{,}649$), followed by patience ($N = 4{,}399$). Lastly, positive and negative reciprocity were added ($N = 2{,}852$). The number of complete observations was much worse for our four measures of revealed preferences. These were, therefore, evaluated one by one in separate specifications (table S9).

Across the models, stated risk willingness was the most strongly associated economic preference ($P < 0.001$ in all specifications, AME up to 4.6 pp per SD; corresponding to a 9.2% size reduction of the unwilling group). It was followed by stated trust, which generally had a small P value but was only suggestively significant after multiple-testing correction (AME up to 2.8 pp per SD). In the first three specifications, we found a rather large but non-significant association with stated altruism (helpful) (AME up to 12.5 pp per SD), but its coefficient was basically zero in the presence of reciprocity (Spearman $\rho = 0.24$ and -0.08 with positive and negative reciprocity, respectively), suggesting conceptual overlap.

We found no significant associations with stated patience or stated altruism (donation). Lastly, stated positive reciprocity was suggestively significant with an effect similar to trust (AME = 2.8 pp per SD), while there was no association with stated negative reciprocity, which makes sense because participation could be considered a positive cooperation without a clear negative action. Likelihood ratio tests suggested that the collection of non-significant stated preferences jointly improved the model fit. Thus, the stated economic preferences as a group appear meaningful for understanding the decision to donate DNA, though the individual effects of each non-risk preference were rather small. Among the economic preferences, stated risk willingness appears to matter the most.

The replication analysis generally supported the above results (table S8b). Stated risk willingness was available for basically the entire SOEP sample ($N = 4{,}125$) and though its coefficient was about half as large, it replicated at the 5% level (AME = 1.9 pp, $P = 0.03$). Stated trust had a comparable AME, but it did not replicate at the 5% level ($P = 0.068$). Likelihood ratio tests for the non-significant stated preferences resulted in small, but non-significant P-values (0.07−0.11). Thus, the replication analysis supported the involvement of stated risk willingness, but also suggests some effect−size heterogeneity between samples or settings. The overall greater willingness in the German data may be an indication that they perceived less risk from donating DNA compared to the Dutch.

Next, none of the revealed economic preferences were associated with the willingness to donate DNA and they correlated only weakly with their stated counterparts (Spearman $\rho = 0.05{-}0.21$; tables S2−3, S9). The highest correlation was observed between stated and revealed altruism as measured by donations (0.21; $N = 1{,}356$). Stated and revealed risk willingness, which had the largest sample overlap ($N = 2{,}371$), correlated at 0.11. Thus, our findings align with recent works that criticize the poor convergence of self-assessed and experimental measures and the limited capacity of the latter to explain real-world behaviour (Frey et al., 2017; Hertwig et al., 2019). Interestingly, a large proportion of respondents rated the questions in the experimental modules as difficult (49.7% to 74.2% of respondents answered ≥ 3 on a 1–5 scale to the question “Was it difficult to answer the questions?”), which is a possible explanation of their poor convergence and performance (Hertwig et al., 2019).

3.2.4. Comparison with “Your DNA, Your Say”

Among the papers based on the “Your DNA, Your Say” survey, our analyses correspond best with the logistic regression results reported for English-speaking countries by Middleton et al. (2019) and for Germany by Voigt et al. (2020). Therefore, we limited the scope of the following comparison of results to these two studies. Importantly, the descriptive evidence we report above in section 3.1 suggest little impact on the study results from the distinction between assaying data donations only vs. donations of both a saliva sample *and* DNA data.

The two previous studies found consistently that younger age, having children and being more religious were associated with greater willingness to donate data to one or more user type. A positive association with tertiary education was found significant in the English-speaking group, but its coefficient was non-significant and close to zero in the previous study on Germany. Both studies found small, non-significant associations with sex (with discordant signs), much similar to the non-significant sex differences we estimated. “Non-white” self-reported ethnicity was associated with lower willingness in the English-speaking countries, but this variable was not investigated by Voigt et al. (2020) because of few observations. Both studies found that familiarity with genetics predicted greater willingness. Lastly, the study by Voigt et al. (2020) found that genetic exceptionalism significantly predicted greater willingness, but this variable was omitted from the study on the English-speaking countries.

Here, in contrast to the previous studies, we found no effect of age in either sample. We also found some evidence of a large effect of non-Western background (AME $= -12.4$ pp), but this association was significant only at the 5% level in some of our specifications in the replication data. Notably, in conflict with the previous studies, we found consistently that being more religious predicted lower willingness.

Moreover, in disagreement with both previous studies and results reported for the UK Biobank (van Alten et al., 2022), we found consistently that higher education predicted lower willingness. To match the previous literature, we tested ad hoc a binary variable for tertiary education, which also produced a negative coefficient. Also, our main analysis suggested a possible inverse U-shaped relationship with educational categories, but the replication analysis did not support a U-shape.

3.2.5. Results for the Personality domain

We proceeded by investigating the Personality domain ($N = 4{,}947$; table S10) and found a positive association with extraversion (AME $= 4.5$ pp per SD) and a suggestive negative association with self-esteem (-1.1 pp per SD). We had not expected a negative association with self-esteem and speculate that higher self-esteem could be important for rejecting the invitation to donate DNA despite the expectation that respondents are, on average, eager to please the researcher and want to contribute to a socially desirable interaction (Grover and Vriens, 2006). In other words, we find it plausible that higher self-esteem makes it easier to say no. The other personality traits generally had small, non-significant coefficients either in the expected direction of effect or close to zero. There was some evidence of a possible weak association with agreeableness and openness, but they were not significant. Likelihood ratio tests suggested the non-significant personality variables could jointly improve the model fit, which supports the notion that dimensions of personality could be involved in the decision to donate DNA but only to a small degree.

Next, to examine the robustness of the stated economic preferences to the inclusion of the Personality domain, we re-ran the four stated preference models with personality included, which reduced the N by about 300−400. The association with stated risk willingness remained strong with a similar coefficient (all $P < 0.001$), while the coefficient of extraversion attenuated by about a quarter to non-significance in the presence of risk willingness. The Pearson correlation between risk willingness and extraversion was 0.27. The smaller coefficients of the other stated economic preference basically did not change. Thus, the coefficients we estimated for the stated preferences and especially risk willingness, were robust to controlling for personality.

3.2.6. Results for the Health & Lifestyle domain

Similar to research on the UK Biobank, our results show that respondents with better self-assessed general health were, on average, more willing to donate DNA for science (table S11). The AME of “Excellent Health” was 11.7 pp relative to “Poor or Moderate Health.” However, among the twenty medical conditions we tested, there were no significant associations. Thus, more research is needed to identify what constituents of general health are responsible. Of note is that the effect of some less common conditions, e.g., Parkinson’s disease (AME = −13.6 pp, $P = 0.233$), were rather large but still non-significant, likely because of lack of power from the limited number of cases (e.g., only 21 cases of Parkinson’s). Also, there was some weak evidence that various health-risk behaviours (drinking, smoking, drug use) and more physical activity were associated with greater willingness to donate DNA.

3.2.7. Results for the Attitudes & Opinions domain

The core surveys include a series of self-assessed items on the confidence in various societal institutions, which we decided to fit separately one by one because of multicollinearity (table S12). In separate regressions on top of the baseline model, more confidence significantly predicted greater willingness. Confidence in science had the largest AME of 4.3 pp per SD, which aligns with the explicit purpose of the genetic data collection described in survey no. 261. Political orientation was considered but excluded because it reduced the sample size and there was no evidence of association.

Next, using some experimental modules, we investigated the opinion of whether the government is ever entitled to store DNA on its entire population and three variables measuring opinions of genetic screening. Because of the smaller sample size of the latter variables, we first evaluated the opinion of the government storing DNA in a separate specification (table S12). This variable was found to have the largest AME among the significant variables: 13.3 and 23.4 pp, respectively, when comparing “only under strict conditions” (58.9% of the sample) and “certainly permissible” (25.5%) versus the base category “absolutely never” (15.6%). Interestingly, aligning with the results above on a possible inverted U-shaped relationship with education, we found that those with primary school and university education more often filled in the answer “absolutely never” compared to the other educational categories. Lastly, the three variables on genetic screening had positive coefficients in the expected direction but were not significant. Thus, we found that confidence in societal institutions and attitudes about the government storing DNA were strong predictors of the willingness to donate DNA for science.

To evaluate the combined effect of our two strongest explanatory variables, i.e., stated risk willingness and opinions about the government storing DNA, we re-estimated the relevant model ad hoc with stated risk willingness included. The coefficient of the opinion variable was basically unchanged. The coefficient of stated risk willingness was slightly larger and

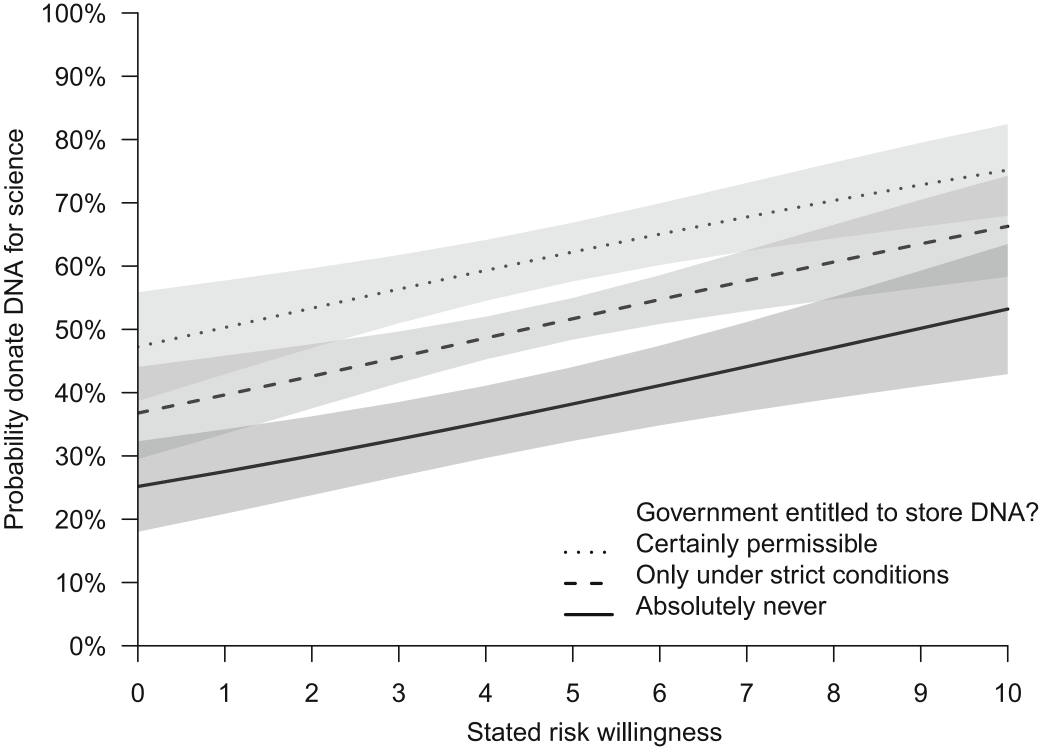

FIGURE 1 The predicted probability of donating DNA for science in the LISS panel
NOTES: The figure shows the probability of donating DNA for science as predicted by a logistic regression model conditional on the range of values of stated risk willingness and the three answer categories to the question "Do you think that the Dutch government is entitled to store DNA information of the entire population?" (and other variables set to their sample means). Grey areas represent 95% confidence bands.

less precise this time (0.122, $SE = 0.03$) but remained within the 95% confidence interval of its model 2b estimate (table 4). Next, to gauge their joint effect in absolute terms, we predicted the probability of donating DNA for the range of values of stated risk willingness and the three categories of the opinion variable (with the other variables fixed at their means) (figure 1). Overall, in combination, these two variables alone predicted a wide range of probabilities from 25.2% to 75.2%, for a total difference of 50 pp.

3.3. Text analysis of stated reasons not to donate DNA

Survey no. 261 provides an interesting opportunity to study stated reasons not to donate DNA because the survey prompted everyone not answering "certainly participate" to provide an open-ended text response to explain ("in a number of words or sentences") what holds them back. While approaches like content and text analysis are not common in economics, they are becoming more important because of the massive growth of online text data and electronic databases (Gentzkow et al., 2019). There are many competing methods with varying complexity. Here, as a descriptive complement to the statistical analysis, we apply a basic word frequency analysis intended to identify commonalities. Thereafter, we present our own interpretation from reading a large number of responses. We acknowledge the more subjective nature of this analysis and that it was designed after the preregistration.

A total of 3,870 respondents had replied different than "certainly participate" and 3,752 of them wrote a text response longer than a single character (e.g., "x" or "-"). We forced the strings to be lowercase and stripped them from non-alphabetical characters. Next, we calculated the number of words per response. The average response was 11 words long (SD = 12.2), with a maximum of 121 words. The first and third quartiles were 3 and 15 words, respectively, and 368 responses were only one word long. Only about 10% of responses were longer than 25 words. Thus, the vast majority of responses were short. There was a small positive correlation between higher willingness and a longer response ($\rho = 0.2$, $P < 0.001$).

We followed standard practice and excluded so-called stop words from the counts (e.g., "the" or "a") by using the online resource "Stopwords" recommended by Gentzkow et al. (2019). The most frequent non-stop word was, unsurprisingly, "DNA," which occurred 840 times across 812 responses (table S13). More interestingly, the second most common word was the word "privacy" (512 times) and variations on the stem "priv" (e.g., "privacy" or "private"), which occurred 679 times in 675 responses. Other related words, such as "personal" were also frequent, with 204 occurrences. The "time" investment was mentioned 166 times. The word stem "secur" occurred 129 times and "hack" occurred 34 times. About 100 responses were variations of "I don't know." 82 responses were about not finding the research interesting. Most other words occurring with high frequency conveyed less meaning without context (e.g., "information"). The word stem "risk" occurred 40 times. Overall, the word frequency analysis found that several privacy-related words were among the words occurring with the highest frequency and that "privacy" was the second most frequent word after "DNA."

Next, we searched the responses specifically for mentioning the financial incentive. There were 53 such responses that we read in full. Noticeably, no responders in the €10 or €20 incentive groups expressed satisfaction with the size of the incentive and instead expressed a shared sentiment that the incentive was too small. However, among the 10 responses about the incentive in the €50 group, only one mentioned slight disappointment about its size. A noteworthy response was "Little is known about the risks and the compensation is small for that uncertainty." Thus, although the effect of the €50 incentive was modest in regression analysis, the text responses at least suggested there was less disappointment expressed in the €50 group. We found it interesting that some respondents stated explicitly that the size of the incentive was not able to compensate for the perceived risks.

Lastly, we read a large number of responses in detail (including the 391 responses longer than 25 words and the 368 one-word responses). Noticeably, many of the shorter responses were highly similar (and many were verbatim) and could therefore be sorted and scanned quickly (e.g., "Fear of [...]") We noticed one response in particular that specifically posed the question of whether respondents would be entitled to compensation for any damages. Insurance discrimination was mentioned as a reason not to donate DNA only 18 times, which was less than expected. In our interpretation, we saw three clear patterns emerging from reading a large number of responses: (i) many respondents expressed concerns about privacy violations and their data falling into the wrong hands, (ii) there were statements expressing genetic exceptionalism (i.e., that genetic data have a special status and must be treated differently than other sensitive medical information) and (iii) uncertainty about whether and how third parties (both public and private) could ever acquire the data. At the same time, many hundreds of respondents simply reconfirmed that they would likely participate despite having replied different than "certainly participate."

4. The expected utility of donating DNA

To stylize our main finding with stated risk preferences and the voiced concerns about privacy risks, we abstracted the decision to donate DNA for science by extending a model on laboratory experiment participation by Abeler and Nosenzo (2015). In this model, people decide whether to participate based on monetary and psycho-social incentives. Our version introduces the possibility of a privacy violation as an element of risk.

Consider the decision to donate DNA to be a rational choice under uncertainty where person i maximizes their subjective utility U_i as a function of wealth (w_i) and other factors (in monetary equivalents). Below, we assume subjective beliefs about the probability of a privacy violation (p_i) and its corresponding subjective loss l_i, with heterogeneous beliefs drawn from their respective population distributions (possibly biased and/or conditional on covariates). For simplicity, we conceptualize various types of privacy violations as a single composite event that could be decomposed further (outside the scope of this study).

We model two appealing factors: the certain monetary reward ($m_i \in \{10, 20, 50\}$); and the psycho-social benefits of donating (b_i), such as virtue signalling or the positive feelings from contributing to public goods (both assumed to be known). The survey did not mention any other direct benefits like genetic counselling, so we omit such a factor. Our model has two deterring factors: (i) the certain direct cost (c_i), such as the time spent or unpleasantness (also assumed to be known), and (ii) the uncertain loss from a privacy violation:

$$E[l_i] = p_i \times l_i + (1 - p_i) \times 0. \tag{4}$$

We assume people make this decision as if their beliefs were true. Because we observed that many panel members had chosen to donate DNA even without an incentive, it appears safe to assume that many people do experience some positive benefit b to compensate for c. Similarly, we observed that many panel members judged the expected loss from a privacy violation to be acceptable, while others were not willing to accept this risk even for a €50 reward.

Thus, the invitation to donate DNA for science is accepted if the expected utility of accepting is greater than the known utility of rejecting (i.e., the status quo, which we assume brings no disutility):

$$E[U(accept)] > U(w), \tag{5}$$

where the expected utility of accepting is a Bernoulli outcome:

$$p \times U(w + m + b - c - l) + (1 - p) \times U(w + m + b - c). \tag{6}$$

Then, the willingness to accept is defined as the size of the monetary reward (m^*) leading to indifference (such that equation 5 holds with equality). If we assume conventional concave utility (i.e., risk aversion) and given b and c, then solving the equation for m identifies the risk premium required to compensate for the perceived risk. Similarly, if no reward is offered, then a risk-averse individual would be willing to donate only if b compensates sufficiently for both c and $E[l]$. The distribution of willingness to accept in the population could be elicited in future research using price lists.

Because most people can be expected to have a limited understanding of the true probability of privacy violations in genetic research data, we expected ambiguity aversion to correlate negatively with the willingness to donate DNA. However, the coefficient of revealed ambiguity aversion was basically zero (table S9).

5. Discussion and conclusion

This preregistered study investigated the willingness to donate DNA for science and its relationship with incentives, economic preferences and socioeconomic characteristics in a large, representative panel of Dutch households. The main result is that, despite guarantees of strict data security and de-identification, a substantial proportion of respondents were hesitant or unwilling to donate DNA. They reported seeing many personal risks due to the sensitive nature of genetic information and the possibility of data breaches and privacy violations. Some respondents also voiced concerns about risks yet unknown to them. Regression analysis showed that proposing a financial incentive as high as €50 was not very effective in swaying respondents to accept these risks, which aligned with text responses explicitly discussing this trade-off in disfavour of the incentive. A few respondents even expressed that the perceived risks are uncertain and cannot be compensated at all.

The first of the two experimental treatments, which presented additional information on the potential benefits and risks of participating, had little to no effect. However, it is possible that the 100-word difference in the information material was not meaningful enough to be detected. Alternatively, the plausibly positive effects of listing more benefits versus the plausibly negative effects of listing more risks could have cancelled out, but we could not test this hypothesis with the available data. If respondents were truly insensitive to this kind of variation in the information material, which future studies would have to confirm, then this knowledge could help simplify the design of the information material used in future data collections and suggests greater comparability across genetic data sets.

The second experimental treatment, which randomized the hypothetical financial incentive, was strongly significant but had only a modest effect. Subsequent economic analysis suggested that, unless the administration and material cost of sending invitations is high (or the sampling frame is limited), it is generally more cost-effective to compensate for the lower acceptance rate of a smaller incentive by simply sending more invitations (or to improve their quality). An alternative to raising the incentive could be to apply "incentive discrimination" in a stratified sampling procedure so that under-represented characteristics get offered higher incentives. However, differential treatment of participants is not always practical or feasible.

Consistent with the risks perceived by the respondents, stated risk willingness was a strong predictor of the willingness to donate DNA. Therefore, we argue that it is inadvisable to offer lottery-based incentives in genetic research, which could intensify the existing self-selection. Moreover, we found associations with trust, positive reciprocity, religiosity, university education, general health, confidence in science or societal institutions and the opinion of whether the government is entitled to store DNA. These estimates could guide the selection of variables to follow up in a causal research design or to include in post-stratification analyses. However, it appears that cheaper stated preference items should be prioritized for this purpose because the more expensive experimentally elicited measures were not associated.

We found an unexpected negative association with university education, which we believe could be explained by this group being, on average, more aware and informed of the ongoing debate about the promises and perils of new genetic technologies and genetic discrimination. They could also be more aware of news about hacks or data leaks, such as those affecting the private genetic testing service 23andMe in the past year (DeGeurin, 2024). Lastly, we found conflicting evidence for certain demographic variables (e.g., age, education and religiosity), possibly because of our representative sample and non-response weighting. To our knowledge, this is the first research on the willingness to donate DNA for science to have been conducted in data not specifically ascertained to study this topic, meaning this is perhaps the first effort to report estimates adjusted for non-response.

For economists interested in genetic research data, our results emphasize the possibility of self-selection on individual differences and that sampling biases should be considered and adjusted for. By opening potential backdoor paths in violation of the exclusion restriction criterion, this bias could be particularly problematic for genetic instrumental variable techniques, which economists have started to adopt (Fletcher, 2018; van Kippersluis and Rietveld, 2018; Biroli et al., 2022). Another insight is that efforts to combat participation biases, e.g., by stratified sampling or post-stratification analysis, could suffice with assaying the cheaper stated preference measures. Lastly, selection on risk willingness could have induced downward bias in genome-wide association studies on risk-taking behaviour, such as Karlsson Linnér et al. (2019), because of the possibility of range censoring of the most risk averse or an increased similarity among self-selected participants. This issue would translate into less powerful genetic risk scores, which is a potential caveat for their growing use in economic research (Becker et al., 2021).

This research has some limitations, the most important being that we studied the stated rather than revealed willingness to donate DNA and that the financial incentive was fictive, though the replication analysis supported the main findings. Nevertheless, the highest incentive condition was still found associated and while its effect size was modest, it effectively reduced complaints about its size in the open-text responses, suggesting respondents did seriously contemplate it as real. Worst case, our estimates establish a lower bound. Secondly, it would have been better if the first experimental treatments were tested as two sub-conditions to distinguish the effect of more information on benefits versus risks. Because the information was presented together, we could not distinguish a truly null effect from a net-zero effect.

Thirdly, despite our large sample, we were under-powered to detect the seemingly large effects of certain variables with few observations, such as people of non-Western background or cases of the critical illnesses (e.g., Parkinson's). It could be worthwhile to follow-up on the large-coefficient variables in larger population samples. Our estimates could inform power analysis of future research. For example, a sample size of $> 20{,}000$ would be required for 80% power to detect the small suggestively significant associations we estimated, e.g., with stated trust. The smaller, non-significant effects we report are probably not a major source of selection bias. Lastly, a large proportion of respondents reported that the questions were difficult to answer in the experimental modules that elicited the economic preferences, which suggests the measures may not be reliable representations of these constructs.

We now propose strategies for minimizing the perceived privacy risks. Most importantly, we recommend reallocating resources from incentives to additional security measures and communication materials on how such measures address specific concerns or threats. At least three approaches exist to improve upon the security measures described in survey no. 261. First, multi-layer access control could be implemented and described, e.g., that raw genetic data will be stored in an encrypted vault available only to a selected few with special security clearance. Raw genetic data pose the greatest risks (especially of re-identification), and special access procedures reduce this threat vector. Second, the raw data could be scrambled and spread across multiple vaults so that the damage from a single breach or break of the encryption is reduced. New blockchain and encryption technologies enable statistical analysis on distributed databases without exposing encrypted data (Chen et al., 2023). An extreme solution could be to implement a type of intelligent banknote neutralization system (IBNS), which would self-destruct the vault upon detecting unauthorized access, or less extremely, to hide watermarks or other such components for traceability. Overall, we think it is worthwhile highlighting to respondents that their raw genetic data are exceptional and, as such, should be treated exceptionally.

Third, most genetically informed research does not require raw genetic data but suffices with genetic summary variables (e.g., genetic relatedness matrices or polygenic scores/indexes) (Harden and Koellinger, 2020). The risk of re-identification is drastically reduced for these data types. Therefore, the data host could preempt requests for raw data by generating harmonized sets of genetic summary variables for distribution under lower-tier security clearance. An example of this solution is the so-called "PGI repository" that was recently created by the Social Science Genetic Association Consortium (SSGAC) (Becker et al., 2021), which provides the research community with a large set of validated polygenic indexes in, e.g., the Health and Retirement Study. While any approved user can access this resource immediately, access to the raw genetic data requires a detailed application and ethical review. Overall, we argue that clever multi-level access procedures that limit the exposure of raw genetic data, in combination with genetic summary variables, are useful strategies for minimizing the perceived risks.

Next, we envision some strategies not directly connected to data security. Importantly, many respondents expressed concern about whether and how the government or law enforcement could get access, which was not addressed explicitly in the information material of survey no. 261. In contrast, the UK Biobank uses the following statement in their invitation material: "Access to the resource by the police or other law enforcement agencies will be acceded to only under court order and UK Biobank will resist such access vigorously in all circumstances" (Bycroft et al., 2018). Thus, it could be useful to preempt these concerns in the invitation, e.g., by clarifying the stance of the host organization or by highlighting existing regulations that prevent such access. Similarly, policymakers could legislate to prevent current and future governments or law enforcement from using the data, e.g., by declaring genetic research data inadmissible as evidence or ban the use of genetic research data for police investigations or insurance risk classification. Our final suggestion is to prioritize resources to purchase liability insurance or a trust fund to pay for compensatory damages, which we believe should be more effective in compensating for the perceived risks than a small flat-rate compensation.

We see several avenues for future research. First, it would be valuable to further survey the risks perceived by the respondents so as to figure out exactly what these risks are and their relative importance to respondents. This knowledge could help decide between competing risk-minimizing measures when funds are limited. Also, little is known about how respondents estimate the damages they expect to experience in the event of a privacy violation. Next, we suggest studying the subjective probabilities of the risk of privacy violations and the willingness to accept of donating DNA, which we believe could be sensitive to recent sensational news about serious data breaches or cold cases being solved thanks to genealogical databases. Prospect theory states that people overweight the probabilities of rare events (Barberis, 2013). Therefore, it would be useful to study whether these subjective probabilities can be influenced by the strategies we propose above, or by other changes to the study design or information material. In particular, it would be useful to investigate whether an insurance scheme would better offset the perceived risks, especially if the willingness to accept turns out to be substantial for many people.

In conclusion, this study provides evidence that research participants consider donating DNA for science to be a risky decision and that the risks involved are not offset by standard incentive amounts offered for research participation. Therefore, addressing the perceived risks of donating DNA should be a key concern for governments and funding agencies that invest in amassing genetic data sets, for policymakers responsible for designing protective legislation and for any host organization involved in genetic data collection. Because we found self-selection on observables, such as risk willingness, we recommend economic researchers working with genetic research data to adjust for sampling biases.

Supporting information

Supplemental material for this article is available at `https://osf.io/32dnr`. The code and instructions for accessing the data that support the findings of this study are available in the Canadian Journal of Economics Dataverse at `https://doi.org/10.5683/SP3/DIYNQA`.

References

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

Barberis, N. C. (2013) "Thirty Years of Prospect Theory in Economics: A Review and Assessment," *Journal of Economic Perspectives* 27, 173–96

Bearth, A., and M. Siegrist (2020) "Psychological factors that determine people's willingness-to-share genetic data for research," *Clinical Genetics* 97, 483–91

Becker, J., C. A. P. Burik, G. Goldman, N. Wang, H. Jayashankar, M. Bennett, D. W. Belsky, R. Karlsson Linnér, R. Ahlskog, A. Kleinman, D. A. Hinds, M. Agee, B. Alipanahi, A. Auton, R. K. Bell, K. Bryc, S. L. Elson, P. Fontanillas, N. A. Furlotte … A. Okbay (2021) "Resource profile and user guide of the Polygenic Index Repository," *Nature Human Behaviour* 5, 1744–58

Benjamin, D.J., D. Cesarini, P. Turley, and A. S. Young (2024) "Social-science genomics: Progress, challenges, and future directions," NBER working paper no. 32404

Biroli, P., T. Galama, H. van Kippersluis, C. A. Rietveld, and K. Thom (2022) "The Economics and econometrics of gene–environment interplay," Tinbergen Institute discussion paper no. 2022-019V

Brañas-Garza, P., D. Jorrat, A. M. Espín, and A. Sánchez (2023) "Paid and hypothetical time preferences are the same: Lab, field and online evidence," *Experimental Economics* 26, 412–34

Bycroft, C., C. Freeman, D. Petkova, G. Band, L. T. Elliott, K. Sharp, A. Motyer, D. Vukcevic, O. Delaneau, J. O'Connell, A. Cortes, S. Welsh, A. Young, M. Effingham, G. McVean, S. Leslie, N. Allen, P. Donnelly, and J. Marchini (2018) "The UK Biobank resource with deep phenotyping and genomic data," *Nature* 562, 203–209

Campos-Mercade, P., A. N. Meier, F. H. Schneider, S. Meier, D. Pope, and E. Wengström (2021) "Monetary incentives increase COVID-19 vaccinations," *Science* 374, 879–82

Chen, L., Y. Mu, L. Zeng, F. Rezaeibagha, and R. H. Deng (2023) "Authenticable data analytics over encrypted data in the cloud," *IEEE Transactions on Information Forensics and Security* 18, 1800–13

Cleave, B.L., N. Nikiforakis, and R. Slonim (2013) "Is there selection bias in laboratory experiments? The case of social and risk preferences," *Experimental Economics* 16, 372–82

DeGeurin, M. (2024, February 15) "*Hackers got nearly 7 million people's data from 23andMe. The firm blamed users in 'very dumb' move*," *The Guardian*

Dimmock, S.G., R. Kouwenberg, and P. P. Wakker (2016) "Ambiguity attitudes in a large representative sample," *Management Science* 62, 1363–80

Drerup, T.H., M. Wibral, and C. Zimpelmann (2023) "Skewness expectations and portfolio choice," *Experimental Economics* 26, 107–44

Falk, A., A. Becker, T. Dohmen, B. Enke, D. Huffman, and U. Sunde (2018) "Global evidence on economic preferences," *Quarterly Journal of Economics* 133, 1645–92

Fletcher, J.M. (2018) "Economics and genetics," *Oxford Research Encyclopedia of Economics and Finance* pp. 1–12

Frey, R., A. Pedroni, R. Mata, J. Rieskamp, and R. Hertwig (2017) "Risk preference shares the psychometric structure of major psychological traits," *Science Advances* 3, e1701381

Fry, A., T. J. Littlejohns, C. Sudlow, N. Doherty, L. Adamska, T. Sprosen, R. Collins, and N. E. Allen (2017) "Comparison of sociodemographic and health-related characteristics of UK Biobank participants with those of the general population," *American Journal of Epidemiology* 186, 1026–34

Gentzkow, M., B. Kelly, and M. Taddy (2019) "Text as data," *Journal of Economic Literature* 57, 535–74

Glynn, P., and P. Greenland (2020) "Contributions of the UK Biobank high impact papers in the era of precision medicine," *European Journal of Epidemiology* 35, 5–10

Grover, R., and M. Vriens (2006) "Response biases in marketing research." In R. Grover and M. Vriens, eds., *The Handbook of Marketing Research*. Thousand Oaks, CA: SAGE Publications, Inc.

Harden, K.P., and P. D. Koellinger (2020) “Using genetics for social science,” *Nature Human Behaviour* 4, 567–76

Harrison, G.W., M. I. Lau, and E. Elisabet Rutström (2009) “Risk attitudes, randomization to treatment, and self-selection into experiments,” *Journal of Economic Behavior & Organization* 70, 498–507

Heckman, J.J. (1979) “Sample selection bias as a specification error,” *Econometrica* 47, 153

Hertwig, R., D. U. Wulff, and R. Mata (2019) “Three gaps and what they may mean for risk preference,” *Philosophical Transactions of the Royal Society B: Biological Sciences* 374, 20180140

Hughes, R.A., N. M. Davies, G. Davey Smith, and K. Tilling (2019) “Selection bias when estimating average treatment effects using one-sample instrumental variable analysis,” *Epidemiology* 30, 350–57

Huynh, L., B. Johns, S.-H. Liu, S. S. Vedula, T. Li, and M. A. Puhan (2014) “Cost-effectiveness of health research study participant recruitment strategies: A systematic review,” *Clinical Trials* 11(5), 576–83

Karlsson Linnér, R., P. Biroli, E. Kong, S. F. W. Meddens, R. Wedow, M. A. Fontana, M. Lebreton, S. P. Tino, A. Abdellaoui, A. R. Hammerschlag, M. G. Nivard, A. Okbay, C. A. Rietveld, P. N. Timshel, M. Trzaskowski, R. d. Vlaming, C. L. Zünd, Y. Bao, L. Buzdugan … J. P. Beauchamp (2019) “Genome-wide association analyses of risk tolerance and risky behaviors in over 1 million individuals identify hundreds of loci and shared genetic influences,” *Nature Genetics* 51, 245–57

Knoef, M., and K. de Vos (2009) “The representativeness of LISS, an online probability panel,” Centerdata technical report

Koellinger, P.D., A. Okbay, H. Kweon, A. Schweinert, R. K. Linnér, J. Goebel, D. Richte, L. Reiber, B. M. Zweck, D. W. Belsky, P. Biroli, R. Mata, E. M. Tucker-Drob, K. P. Harden, G. Wagner, and R. Hertwig (2023) “Cohort profile: Genetic data in the German Socio-Economic Panel Innovation sample (SOEP-G),” *PLOS ONE* 18, e0294896

Middleton, A., R. Milne, M. A. Almarri, S. Anwer, J. Atutornu, E. E. Baranova, P. Bevan, M. Cerezo, Y. Cong, C. Critchley, J. Fernow, P. Goodhand, Q. Hasan, A. Hibino, G. Houeland, H. C. Howard, S. Z. Hussain, C. I. Malmgren, V. L. Izhevskaya … K. I. Morley (2020) “Global public perceptions of genomic data sharing: What shapes the willingness to donate DNA and health data?,” *American Journal of Human Genetics* 107, 743–52

Middleton, A., R. Milne, A. Thorogood, E. Kleiderman, E. Niemiec, B. Prainsack, L. Farley, P. Bevan, C. Steed, J. Smith, D. Vears, J. Atutornu, H. C. Howard, and K. I. Morley (2019) “Attitudes of publics who are unwilling to donate DNA data for research,” *European Journal of Medical Genetics* 62, 316–23

Middleton, A., E. Niemiec, B. Prainsack, J. Bobe, L. Farley, C. Steed, J. Smith, P. Bevan, N. Bonhomme, E. Kleiderman, A. Thorogood, C. Schickhardt, C. Garattini, D. Vears, K. Littler, N. Banner, E. Scott, N. V. Kovalevskaya, E. Levin, K. I. Morley, and H. C. Howard (2018) “’Your DNA, Your Say’: Global survey gathering attitudes toward genomics: Design, delivery and methods,” *Personalized Medicine* 15, 311–18

Milne, R., K. I. Morley, H. Howard, E. Niemiec, D. Nicol, C. Critchley, B. Prainsack, D. Vears, J. Smith, C. Steed, P. Bevan, J. Atutornu, L. Farley, P. Goodhand, A. Thorogood, E. Kleiderman, and A. Middleton (2019) “Trust in genomic data sharing among members of the general public in the UK, USA, Canada and Australia,” *Human Genetics* 138, 1237–46

Munafò, M.R., K. Tilling, A. E. Taylor, D. M. Evans, and G. D. Smith (2018) “Collider scope: When selection bias can substantially influence observed associations,” *International Journal of Epidemiology* 47, 226–35

Noussair, C.N., S. T. Trautmann, and G. van de Kuilen (2014) “Higher order risk attitudes, demographics, and financial decisions,” *Review of Economic Studies* 81, 325–55

O’Connor, L.J. (2021) “The distribution of common-variant effect sizes,” *Nature Genetics* 53, 1243–49

Pirastu, N., M. Cordioli, P. Nandakumar, G. Mignogna, A. Abdellaoui, B. Hollis, M. Kanai, V. M. Rajagopal, P. D. B. Parolo, N. Baya, C. E. Carey, J. Karjalainen, T. D. Als, M. D.

van der Zee, F. R. Day, K. K. Ong, FinnGen Study, 23andMe Research Team, M. Agee … A. Ganna (2021) "Genetic analyses identify widespread sex-differential participation bias," *Nature Genetics* 53, 663–71
Richter, D., and J. Schupp (2015) "The SOEP Innovation Sample (SOEP IS)," *Schmollers Jahrbuch* 135, 389–99
Roe, B.E., T. C. Haab, D. Q. Beversdorf, H. H. Gu, and M. R. Tilley (2009) "Risk-attitude selection bias in subject pools for experiments involving neuroimaging and blood samples," *Journal of Economic Psychology* 30, 181–89
Scherpenzeel, A. (2011) "Data collection in a probability-based internet panel: How the LISS Panel was built and how it can be used," *Bulletin of Sociological Methodology* 109, 56–61
Scherpenzeel, A., and M. Das (2010) "'True' longitudinal and probability-based internet panels: Evidence from the Netherlands." In M. Das, P. Ester, and L. Kaczmirek, eds., *Social and Behavioral Research and the Internet: Advances in Applied Methods and Research Strategies*., pp. 77–104. Boca Raton, FL: Taylor & Francis
Schoeler, T., D. Speed, E. Porcu, N. Pirastu, J.-B. Pingault, and Z. Kutalik (2023) "Participation bias in the UK Biobank distorts genetic associations and downstream analyses," *Nature Human Behaviour* 7(7), 1216–27
Slonim, R., C. Wang, E. Garbarino, and D. Merrett (2013) "Opting-in: Participation bias in economic experiments," *Journal of Economic Behavior & Organization* 90, 43–70
The All of Us Program (2019) "The 'All of Us' Research Program," *New England Journal of Medicine* 381, 668–76
Thiemann, P., J. Schulz, U. Sunde, and C. Thöni (2022) "Selection into experiments: New evidence on the role of preferences, cognition, and recruitment protocols," *Journal of Behavioral and Experimental Economics* 98, 101871
Thomas, C., S. Mulnick, N. Krucien, and K. Marsh (2022) "How do study design features and participant characteristics influence willingness to participate in clinical trials? Results from a choice experiment," *BMC Medical Research Methodology* 22, 323
van Alten, S., B. W. Domingue, J. Faul, T. Galama, and A. T. Marees (2022) "Correcting for volunteer bias in GWAS uncovers novel genetic variants and increases heritability estimates," MedRxiv preprint. https://doi.org/10.1101/2022.11.10.22282137
——— (2023) "Should representativeness be avoided? Reweighting the UK Biobank corrects for pervasive selection bias due to volunteering," CESR-Schaeffer working paper no. 2023-002
van Kippersluis, H., and C. A. Rietveld (2018) "Beyond plausibly exogenous," *Econometrics Journal* 21, 316–31
Voigt, T.H., V. Holtz, E. Niemiec, H. C. Howard, A. Middleton, and B. Prainsack (2020) "Willingness to donate genomic and other medical data: Results from Germany," *European Journal of Human Genetics* 28, 1000–1009
Walter, J., and M. David (2016) "Who's willing? Characteristics associated with willingness to participate in clinical research," *IRB: Ethics & Human Research* 38(2), 15–20
Wang, Y., S. Namba, E. Lopera, S. Kerminen, K. Tsuo, K. Läll, M. Kanai, W. Zhou, K.-H. Wu, M.-J. Favé, L. Bhatta, P. Awadalla, B. Brumpton, P. Deelen, K. Hveem, V. Lo Faro, R. Mägi, Y. Murakami, S. Sanna … B. M. Neale (2023) "Global Biobank analyses provide lessons for developing polygenic risk scores across diverse cohorts," *Cell Genomics* 3, 100241
Zeggini, E., A. L. Gloyn, A. C. Barton, and L. V. Wain (2019) "Translational genomics and precision medicine: Moving from the lab to the clinic," *Science* 365, 1409–13

Appendix A1: Extended acknowledgements

This research was reviewed and approved by the Committee Ethics and Data of the Leiden Law School of Leiden University (application number 2023-24). This publication is part of the project "The economic consequences of new genetic testing technologies" with project number Vl.Veni.221E.080 of the NWO Talent Programme Veni – Social Sciences and Humanities, which is (partly) financed by the Dutch Research Council (NWO).

The LISS panel data were collected by Centerdata (Tilburg University, the Netherlands) through its MESS project funded by the Netherlands Organization for Scientific Research (Scherpenzeel and Das, 2010). The panel members provided informed consent. We are grateful to the LISS panel members and the affiliated researchers for making this study possible. The LISS data were accessed under application number(s) 9717 (R.K.L) and 12384 (M.J). The de-identified quotes reproduced in the article were granted written permission from Centerdata.

The SOEP panel data were collected by DIW Berlin (Richter and Schupp, 2015). The panel members provided informed consent. We thank the SOEP-IS panel members and the affiliated researchers for making this study possible. The SOEP-IS data[7] were accessed under application number 6689 (R.K.L).

The authors declare no conflicts of interest.

Appendix A2: Author contributions

The study was conceived by R.K.L. and designed by M.J. and R.K.L. R.K.L. was the principal investigator responsible for the project and its administration. M.J. was responsible for the data preparation and statistical analysis under the guidance of R.K.L. The replication analysis was done by R.K.L. R.K.L. drafted the manuscript with support from M.J. R.K.L. and M.J. reviewed and edited the draft manuscript.

Appendix A3: Details on the variable selection and preregistration

For the preregistration, we scanned all core and non-core survey modules to find candidate variables for consideration, focusing on stated and revealed economic preferences. Because of their relevance to the social sciences and policy research, we also selected many variables that we categorized as from the Demography, Personality, Health & Lifestyle or Attitudes & Opinions domains. To reduce the researcher's degrees of freedom, we preregistered these variables based on the survey descriptions and codebooks without looking at the underlying data. Therefore, a series of decision rules and procedures were preregistered to control how we would later prioritize among potentially correlated or conceptually overlapping measures. The rules stated that variables were to be prioritized by: (i) sample size, (ii) variable precision (e.g., continuous over binary) and (iii) representativeness of definitions in previous research. The preregistration listed a curated selection of about 75 candidate variables (tables S1–3).

Further, we preregistered the plan to prioritize incentivized measures over stated counterparts whenever the former was available at a comparable sample size. We later found this was never the case (t) and per the protocol, incentivized measures were therefore set aside and analyzed separately in auxiliary models. Unfortunately, we also found it infeasible to fit most or all candidate variables at once because of considerable non-overlapping missingness (remaining $N \sim 250$). Therefore, we adapted the protocol during the study to instead define a baseline model with little missingness, on which we evaluated sets of variables that are conceptually meaningful to fit together while retaining a satisfactory sample size.

Appendix A4: Details on the preregistered hypotheses

The preregistration listed a few statistical hypotheses to be tested with logistic regression analysis, most of which were interaction terms between the experimental treatments and

7 DOI: https://doi.org/10.5684/soep.is.2021.

selected variables of special interest (e.g., risk willingness). First, it was hypothesized that being randomized to receive more information in experimental treatments (i) would, on average, increase the willingness to donate DNA for science. The reasoning was that more information should reduce uncertainty and even though benefits and risks were presented in tandem, previous research has shown that participants in clinical trials tend to focus on the benefits (Bearth and Siegrist, 2020; Thomas et al., 2022). Secondly, it was hypothesized based on expected utility theory that higher financial incentives would increase the overall willingness.

With respect to the candidate explanatory variables, only two hypotheses were preregistered that were not interactions with the experimental treatments: (i) because we lacked exogenous variation in the non-experimental variables, we preregistered that any statistically significant associations would be considered correlational and not causal and (ii) that exogenous variation in years of schooling should increase the willingness causally. The preregistration proposed a quasi-experimental analysis using a Dutch schooling reform, which was later abandoned because of insufficient observations. Thus, the causal effect of education was not tested. Lastly, a series of interaction terms with the experimental treatments were preregistered, e.g., that receiving more information should have a comparably larger effect among people averse to ambiguity, or that higher financial incentive should have a larger effect among people with lower income. As we report in section 3, none of the preregistered interactions were statistically significant and an ex post power analysis showed insufficient power to detect reasonable effect size ranges. For more information on the tested interaction terms, please see the preregistration.[8]

Appendix A5: Details on the representativeness of the LISS

Centerdata allocates many resources for better response rates and lower attrition, e.g., remuneration for survey time (today, €15 per hour), or a free computer with Internet access (for households without). Non-responders get contacted repeatedly via multiple channels for encouragement or to confirm a wish to withdraw. The response rate of the annual core surveys sent to all active panel members is stable at about 80% to 90%, while the sampling frame and response rates of individual experimental modules can vary (e.g., incentivized experiments are rarely deployed in the full panel because of the higher costs involved). The response rate to non-core modules is relatively high (70% to 80%) and the overall attrition rate each year is low (∼10%).

Overcompensating refreshment samples, which also oversample underrepresented groups, have been drawn periodically every two or three years. Although the response rate of some groups, e.g., the oldest or minority backgrounds (Knoef and de Vos, 2009) is still lower than expected, the differences are small enough for the panel to be cited as being representative (Drerup et al., 2023). Therefore, the dataset does not include any sample weights and post-stratification is rarely seen applied to these data. Nevertheless, in the next paragraph, we confirm the representativeness of the general Dutch population. Furthermore, all regressions were inverse-probability weighted to adjust for the non-response to survey no. 261.

To evaluate sample representativeness, we compared our sampling frame to the Dutch population on a set of key demographic indicators pulled from Statistics Netherlands (table S5). There were no substantial differences, but we noticed slight oversampling of women, ages 65 to 80 (while ages 80+ were undersampled), Native background and

8 The preregistered analysis plan is available at `https://osf.io/8tzp9`.

Western background. There was also some oversampling of the highest educational category. However, this discrepancy could come from the way Statistics Netherlands groups certain higher vocational degrees as intermediate. Overall, we consider our sampling frame to be largely representative, so we did not conduct any post-stratification or weighting to match the general population.

Appendix A6: Details on the imputation of missing values

Despite the cross-sectional analysis due to survey no. 261 being assayed only once, we exploited the longitudinal data to: (a) add explanatory variables not measured in the 2020 wave (e.g., on economic preferences), (b) improve data coverage by replacing missing values in the core surveys with the closest preceding wave (at most one previous wave) and (c) reduce measurement error by averaging repeated measures of continuous variables. For the core surveys, we averaged the 2020 wave with at most two preceding waves. For the stated economic preferences, we averaged repeated measures whenever available. Before averaging, we first checked that the means and variances were similar and that correlations were non-zero or higher (tables S2–S4). The imputation of missing values often boosted the sample size by many hundreds of observations.

Appendix A7: Details on the stated economic preferences

We defined the following seven measures of stated economic preferences:

1. Patience ($N = 4{,}728$)
2. Risk willingness ($N = 5{,}022$)
3. and 4. Positive and negative reciprocity ($N = 3{,}078$)
5. Altruism (donation; $N = 5{,}304$)
6. Altruism (helpful; $N = 5{,}245$)
7. Trust ($N = 5{,}246$)

The following definitions correspond well with the Global Preference Survey (Falk et al., 2018).

Stated patience was assayed on a 0–10 scale by asking the question "How willing are you by nature to forego something today, if you stand to benefit from that at some point in the future?" (table S3). We averaged two repeated measures ($r \sim 0.34$), for a total N of 4,728. The final measure had a mean of 6.73 (SD = 1.88).

Stated risk willingness (max $N = 5{,}022$) was averaged over nine repeated measures on a 0–10 scale of slight variations of the question "Are you, in general, someone who is willing to take risks or someone who avoids risks?" (table S2 reports the exact wording of each question). We excluded two questions that were instead measured on a 0–7 scale. The means and standard deviations among the averaged questions were comparable (table S2) and the pairwise correlations were all in the range of 0.31–0.55 (table S4), without any noticeable outliers. The final measure had a mean of 4.78 (SD = 1.83).

Stated positive reciprocity (max $N = 3{,}078$) was defined by a single measure as the sum over three questions (on a 1–5 scale, shifted to 0–4): "If someone does me a favour, I am willing to do something in return," "I will do my very best to help someone who once helped me in the past" and "I am willing to make an effort to help someone who helped me in the past" (table S3). Stated negative reciprocity (max $N = 3{,}078$) was defined analogously with the three questions: "If I am treated very unjustly, then I will do whatever it takes to have my revenge," "If someone puts me in a difficult position, then I will do the same to him or her" and "If someone insults me," then I shall repay that person in kind" (table S3). The final scores ranged from 0 to 12, with means of 9.39 (SD = 2.16) and 4.83 (SD = 2.84), respectively.

We defined two complementary measures of stated altruism (donation and helpful): (i) having donated to an organization "for humanitarian aid or human rights" and/or "environmental protection, peace organization or animal rights organization" in the last 12 months (three repeated measures, merged as "yes" for any positive answer), max $N = 5{,}304$), and (ii) the self-assessed rating of being helpful as an important guiding principle in life on a 1–7 scale, averaged over three repeated measures (Pearson $r \sim 0.45$, max $N = 5{,}245$). About 28.5% of respondents answered "yes" to the donating question at least once and the mean of altruism (helpful) was 5.94 (SD = 0.82) (table S3).

Stated trust (in people) was assayed on a 0–10 scale with the question "Generally speaking, would you say that most people can be trusted, or that you can't be too careful in dealing with people?" and was averaged over three repeated measurements ($r \sim 0.72$, max $N = 5{,}246$) (table S3), resulting in a mean of 6.06 (SD = 2.01).

Appendix A8: Details on the revealed economic preferences

Here we provide additional details on how we defined our four measures of revealed economic preferences:

1 Risk willingness ($N = 2{,}374$)
2 Ambiguity aversion ($N = 772$)
3 Altruism (donation) ($N = 1{,}360$)
4 Trust ($N = 557$)

Incentivized economic experiments are not part of the core surveys but have been elicited over the years as experimental modules. Unfortunately, experimental modules are rarely or never deployed in the full panel, nor are they repeated over time. Also, a recurring strategy has been to incentivize only a subset of games or participants and, after that, to demonstrate or argue for little differences between real and hypothetical conditions (Noussair et al., 2014). Nevertheless, for the preregistration, we scanned the survey descriptions and identified six modules eliciting risk preferences (two were later excluded, see below); and one module each on ambiguity preferences, altruism (donation) and trust (in people); that we preregistered for consideration. The interested reader can find the relevant code books in the data archive of the LISS panel (`https://dataarchive.lissdata.nl`) by searching for the survey numbers we report below. The following paragraphs describe how we used these experiments to code the revealed preferences analyzed in this study.

Appendix A9: Measuring revealed risk preferences

Incentivized measures of risk preferences were found in the surveys numbers 38, 44, 81, 135 (part 2), 153 and 166 (table S2). The first four modules have been analyzed in published studies (see table S2), while to our knowledge, the latter two have not. The sample overlap of each module with survey no. 261 was limited (covering only 1.8% to 25.7% of the 5,361 respondents). Because of the limited overlap and the fact that these modules employed some type of staircase procedure with binary lottery choices (providing a clear rank order of more risk-taking), we decided to adapt the study protocol to create a single composite measure rather than to pursue up to six competing versions with large differences in sample size. For the reasons explained below, the surveys no. 81 and 166 were excluded from the composite. We first discuss the design of the composite before describing the key details of each experiment.

To create the composite measure, we first converted the rank order revealed by each experiment to a percentile rank, which was then averaged over available non-missing observations, the number of which varied across respondents ($N = 2{,}374$). Ultimately, 43.6% of the observations in the composite was based on a single module, 26.3% on two modules, 18.2% on three modules and 11.9% were observed in all four. The percentile ranks were correlated among themselves (pairwise Spearman $\rho = 0.09{-}0.22$) as well as with stated risk willingness: pairwise $\rho = 0.05{-}0.19$ ($N = 579{-}1{,}377$). Importantly, the composite measure provided about 1,000 extra observations and correlated significantly with stated risk willingness in the expected direction and not much worse than the best individual measure ($\rho \sim 0.13$ vs 0.19). Also, as expected, it correlated negatively with being female ($\rho = -0.13$ vs. $\rho = -0.21$ for stated risk willingness). In robustness checks, we replaced the composite with each of the individual measures, which all produced null results consistent with the findings of the main analysis that instead used the composite (see section 3).

Survey no. 38 was the first experimental survey to elicit incentivized risk preferences in the LISS ($N = 1{,}377$ overlapping with no. 261). The reward conditions were randomized so that 40% of the players were in the paid condition, of which 10% would be selected at random for actual payment, meaning that only 4% of the players were actually rewarded a prize. The players knew from the start whether they were in the paid condition or not but were told only at the end if they had been selected for payment. The authors of the original study of this module concluded that "there are no significant differences between the Real and the Hypo treatments for any of the measures" and "the result supports the view that hypothetical lottery questions are a valid, unbiased, instrument to elicit risk attitudes on survey panels where real financial incentives cannot easily be implemented" (Noussair et al., 2014). The experiment consisted of four parts, of which one was designed to measure "first-order" risk preferences. The other parts were designed to measure higher-order risk preference parameters (prudence and temperance), which we do not study here.

In the game from survey no. 38 used in the composite, subjects were presented with a staircase procedure of five lotteries, in which they had to choose between a fixed risky payoff (50% of €5 and 50% of €65; expected value = €35) versus a safe payoff with increasing value (i.e., 20, 25, 30, 35 and, finally, 40). For the composite measure, we computed the rank order of switching from the risky to the safe bet (or ending risky), which thus had six levels. The ρ between this rank order and stated risk willingness was 0.173 and -0.132 with being female, suggesting it successfully tagged variation in risk willingness. As a side note, the original study instead analyzed this game as a count variable of the number of safe choices, which we tried ad hoc but found to be less strongly correlated with stated risk willingness (~ 0.154). Thus, we believe our procedure improved upon the original study. The players found the game difficult given that 65.9% of the players answered ≥ 3 on a 1−5 scale to the question "Was it difficult to answer the questions?"

Survey no. 44 was focused on eliciting ambiguity aversion (see below), but also elicited risk preferences with two games ($N = 780$ overlapping with no. 261). The reward conditions were randomized so that 50% of the players could win a prize based on their choice in one of the rounds played to elicit ambiguity but not risk preferences. The players knew from the start whether they were in the paid condition or not, but not which game would be picked. The players played two games eliciting risk preferences that differed by presenting fairly large (expected value of fixed risky choice = €500) versus extra large amounts (expected value of fixed risky choice =€9,000). In both cases, the games were a staircase procedure where the players decided between a fixed risky payoff (50% of €X and 50% of €Y) versus a safe payoff with an increasing or decreasing value depending on the player picking safe or risky. The game iterated until the players indicated indifference, or until the change of the safe value was less than 100 euros. The game with the smaller amounts converged faster and provided fifteen end states, while the game with larger amounts provided 103 end states, both providing a rank order of more risk willingness. The ρ between the two games was 0.59 and their respective ρ with stated risk willingness were positive but non-significant: 0.0539 and 0.045 (both $P > 0.05$). As expected, the two ranks both correlated negatively with being female (-0.041 and -0.072). The players found the game difficult given that 57.7% of the players answered ≥ 3 on a 1–5 scale to the question “Was it difficult to answer the questions?”

Survey no. 81 was focused on eliciting a large number of competing risk preference measures. This meant that the experimental procedure was divided into many smaller subgroups with imperfect overlap. Therefore, the procedure most similar to the other modules on risk preference had unsatisfactory overlap with survey no. 261 (only 98 respondents). Because of the very limited overlap, we decided not to include this measure in the composite. Also, this module was rated as the most difficult by the respondents: 70.1% reported $\geq$ 3 on a 1–5 scale to the question “Was it difficult to answer the questions?”

Survey no. 135 was focused on playing a variety of incentivized stock market games and no. 135 part 2 elicited risk preferences ($N = 1{,}239$ overlapping with no. 261) and has been studied by Drerup et al. (2023). The other parts were incentivized, but not the part 2 module. The game was a staircase procedure where the players decided between a fixed risky payoff (50% of €0 and 50% of €300 with an expected value of 150) versus a safe payoff with an increasing or decreasing value depending on the player picking the safe or risky option. The procedure provided 32 end states with a consistent rank order of more risk willingness. The ρ with stated risk willingness was the largest among these modules: 0.1895 and it correlated negatively with being female (-0.113). The players found the game difficult given that 55.3% of the players answered $\geq$ 3 on a 1–5 scale to the question “Was it difficult to answer the questions?”

Survey no. 153 was designed somewhat differently by: (i) the decision being between two risky lotteries (and not the conventional safe choice) and (ii) by simultaneously varying the variance of one of the two lotteries ($N = 579$ overlapping with no. 261). The exact logic behind the design was difficult to confirm because this module has yet to appear in published research (or perhaps the experiment did not work out as intended). Participants did not know at the start whether they were among the 10% to be chosen at random for payment according to one of their choices. The game was a staircase procedure where the players decided between two risky lotteries (always with 50/50 probabilities), one with fixed rewards (50% of €50 and 50% of €150, expected value = 100) and one with an increasing expected value (going from 75 to 100) and decreasing variance (going towards zero), until the varying lottery collapsed into a safe bet with the same value as the expected value of the fixed risky lottery. Here, we employ the logic that the fixed risky lottery starts with both a higher expected value and a higher maximum payoff, making this lottery more attractive if we ignore the fact that the variance of the alternative lottery was designed to always be smaller. In this way, we generated a consistent rank order with 26 end states based on the switching from the high expected value and high variance lottery to the second lottery with: (a) smaller but increasing expected value (converging to the expected value of the high variance lottery) and (b) lower and further decreasing variance (collapsing to zero). The correlation with stated risk willingness was positive and significant at the 5% level ($\rho = 0.101$) and -0.094 with being female. The players found this game the least difficult given that 49.7% of the players answered $\geq$ 3 on a 1–5 scale to the question “Was it difficult to answer the questions?”

Lastly, the most recent experimental module eliciting risk preferences was survey no. 166 ($N = 761$ overlapping with no. 261). Here, respondents knew that 5% of the sample would be picked at random to get one of their choices paid out. Risk preferences were elicited with a single question that mimicked a staircase procedure between one risky option with fixed payoffs (50% chance of €15 and 50% chance of €90 with an expected value of 52.5) and one safe option with decreasing value (going from 90 to 10 in steps of 5). However, the respondents were presented with all options at once and were asked to select the row where they felt motivated to switch from the safe options to the risky option. Although this procedure was fairly simple and comparable to the above procedures, this rank order did not correlate significantly with stated risk willingness ($\rho = 0.013$) nor with being female ($\rho = 0.001$). Therefore, to avoid diluting the composite measure, we decided not to include it in the composite. The players found the game difficult given that 66.3% of the players answered ≥ 3 on a 1–5 scale to the question "Was it difficult to answer the questions?"

Appendix A10: Measuring revealed ambiguity aversion

To define ambiguity aversion, we sourced data from survey no. 44 ($N = 772$ overlap with no. 261). The reward conditions were randomized so that 50% of the players could win a prize based on their choice in one of the rounds played. The players knew from the start whether they were in the paid condition or not, but not which game would be picked. The authors of the original study of this experiment did, however, find some differences between hypothetical and real conditions (Dimmock et al., 2016). To maximize the sample size and avoid mixing different games, we did not distinguish between hypothetical and real conditions. The relevant game consisted of a staircase procedure of picking between two competing binary lotteries of which one had known and the other had unknown probabilities. The lottery with known probabilities was made less and less attractive until players felt it was motivated to switch to the lottery with unknown probabilities or stated they were indifferent. This procedure provided a consistent rank order of ambiguity aversion with 62 possible end states (coded so that higher value measured greater aversion to ambiguity), which we used as our measure of revealed ambiguity aversion. The original study did not find much noteworthy overlap with real-word traits and ambiguity aversion (Dimmock et al., 2016), similar to our null results for this variable (see section 3). The players found the game difficult given that 57.7% of the players answered ≥ 3 on a 1–5 scale to the question "Was it difficult to answer the questions?"

Appendix A11: Measuring revealed altruism

We sourced our measure of revealed altruism (donation) from survey no. 310 (Greener than Others) ($N = 1{,}360$ overlap with no. 261). The aim of the survey was to assay decisions about donating money to charitable environmental organizations, which is similar to our stated altruism (donation) measure (see table 1). Respondents had a 1 in 30 probability of winning 40 euros in a lottery and were asked, if they were to win the lottery, how much they wished to donate (0, 10, 20, 30 or the full 40 euros) to an environmental organization. We use this variable to construct a binary measure of altruism (donation), with those who responded that they would be willing to donate a non-zero amount (i.e., 10, 20, 30 or 40 euros) coded as one and those who stated that they would donate zero euros on winning the lottery coded as zero. This binary measure correlated at $\rho = 0.213$ with stated altruism (donation) and $\rho = 0.05$ with stated altruism (helpful). The players found this game the least difficult among all the experimental modules we sourced: only 44.4% of the players answered ≥ 3 on a 1–5 scale to the question "Was it difficult to answer the questions?"

Appendix A12: Measuring revealed trust

The source for the measure of revealed trust is survey no. 77 (Betrayal, Beliefs and the Demographics of Social Capital). All respondents were incentivized and knew that one game would be picked at random for the reward. To generate a measure of trust, we used two questions that measured subjective belief's about the proportion of other players that would betray in a later one-shot betrayal game (that was played with only a smaller subset of respondents). The first question asked one group of players to directly estimate the proportion (in steps of 10) of potential second-movers (picking between cooperate or betray) that were going to cooperate rather than betray in a one-shot trust game. The proportion provides a consistent percentile rank order of having more trust in that a larger proportion of second-movers will cooperate rather than betray.

The second question asked another non-overlapping group of players to instead place a bet on the proportion of potential second-movers (picking between cooperate or betray) that were going to cooperate rather than betray. The bets elicited basically the same belief about what proportion of second-movers were going to betray. Since the samples in which these two forms of this question were asked are mutually exclusive, we create a composite measure ($N = 557$ overlap with no. 261) of revealed trust by combining the rank orders obtained from the two formats of this question. The players found this game the most difficult among all the experimental modules we sourced given that 74.2% of the players answered ≥ 3 on a 1−5 scale to the question "Was it difficult to answer the questions?" Notably, more than half of the players could not figure out the correct payoffs in the test questions prior to the actual games.

Appendix A13: Details on the German SOEP replication data

The SOEP-IS members answer annually a shorter version of the extensive core questionnaire. Also, various experimental modules are deployed each year, but these are seldom or never repeated and have limited overlap with the 2019 wave. Ultimately, we were able to code most baseline covariates. Furthermore, their measure of stated risk willingness was essentially the same as that in LISS (see above) and comparable measures were found for trust, positive and negative reciprocity, patience and altruism (defined as the stated amount of donation to a good cause in case of a windfall gain of €1,000). We did not pursue replication of our model specifications in the Revealed Economic Preferences, Personality, Health & Lifestyle or Attitudes & Opinions domains either because these domains were mostly not associated in our main analysis (tables S10−12) or because there were no clearly overlapping variables (or proxies) available in the replication data.