Impact and Responsibility in Giving Behavior: A Replication and Extension

Ting-Yi Lin[1], Dennie van Dolder[2], Martijn J. van den Assem[3], Alexander G. Genevsky[1], Ale Smidts[1], and Maarten A. S. Boksem[1]

[1] Rotterdam School of Management, Erasmus University Rotterdam
[2] Department of Economics, University of Essex
[3] School of Business and Economics, Vrije Universiteit (VU) Amsterdam

Abstract: Do donors prioritize utilitarian considerations and mostly care about how much a donation improves well-being ("impact"), or do they prioritize equity-and-desert considerations and mostly care about whether recipients are to blame for their predicaments ("responsibility")? Eberhardt Hiabu et al. (2026) provide the first direct empirical comparison of these two motives behind giving behavior, drawing on data from a television show in which audience members allocate €10,000 among three candidates who face financial difficulties, combined with structured evaluations by independent raters. They find that perceived well-being impact plays a larger role in donation decisions than perceived personal responsibility. The present paper replicates and extends that study. First, we re-analyze the audience donations, employing more raters per case and a simplified impact measure. Second, we additionally elicit raters' individual willingness to donate, and analyze how these stated propensities are related to raters' own perceptions of impact and responsibility. In both analyses, impact is a stronger predictor of giving than responsibility, and dominance analysis confirms its greater explanatory power. Furthermore, we find that the patterns for male and female raters' willingness to donate are similar. Altogether, the results offer robust evidence that both utilitarian and equity-and-desert-based considerations influence donation decisions, and that utilitarian thinking tends to dominate.

JEL: D63, D64, D91

Keywords: giving, utilitarianism, equity, desert, impact, personal control, responsibility, replication

Version: April 2026

1. Introduction

In moral philosophy, utilitarianism and theories of equity and desert represent two distinct paradigms for thinking about giving behavior. Utilitarianism advocates maximizing societal well-being, and exclusively considers the outcomes of actions, regardless of how these are achieved. In contrast, theories of equity and desert maintain that outcomes should reflect individuals' roles in the process leading to these outcomes. These paradigms imply different donor motivations: a utilitarian donor prioritizes how much a donation improves well-being (impact), while an equity-and-desert-oriented donor focuses on whether the recipient is to blame (responsibility) for their predicament. While various empirical studies find that donors are sensitive to the well-being impact of their donations, others find that such considerations are often neglected. In contrast, there is robust evidence that perceptions of personal responsibility strongly affect generosity.[1]

Eberhardt Hiabu et al. (2026; henceforth: ESVV) provide the first direct empirical comparison of these two motives. They study giving decisions in the Dutch TV show *Geld Maakt Gelukkig* ("Money Buys Happiness"), where one hundred audience members allocate €10,000 among three candidates who face a problem beyond their financial capacity. ESVV link actual audience donation outcomes to structured evaluations by independent raters, who assessed each candidate on (i) the expected well-being impact of a donation and (ii) the candidate's personal control over the cause of their predicament. In contrast with the picture emerging from the literature, they find that impact outweighs responsibility: giving increases more with the perceived well-being benefit of donating than it decreases with the perceived control of the candidates over the cause of their situation, and impact accounts for more than twice the explanatory power of personal control.

The present study replicates and extends the findings of ESVV. First, we replicate their core analyses for a subset of the original *Geld Maakt Gelukkig* episodes, employing both a larger number of raters per case to improve the reliability of the measures, and a simplified impact measure to reduce raters' cognitive load. Second, we extend their analysis by linking individual raters' stated willingness to donate to their own perceptions of impact and personal control. The results of the two analyses closely mirror those of ESVV: giving increases more with perceived impact than it decreases with personal control, and the contribution of impact to the empirical fit is substantially greater than that of personal control. Furthermore, we find similar patterns when we separately analyze how male and female raters' willingness to donate is driven by impact and personal control.

[1] For a literature review, see Eberhardt Hiabu et al. (2026).

In sum, we confirm the robustness of ESVV's findings. Across distinct data sets, different impact measures, and both group and individual-level decisions, we find that donation decisions are influenced by both utilitarian and equity-and-desert-based considerations. Utilitarian thinking, however, tends to dominate.

2. Description of ESVV

ESVV use the Dutch TV show *Geld Maakt Gelukkig* (henceforth: GMG), which aired in 2014-2015. In GMG, three candidates present their troubles to an audience of one hundred people. Each audience member is given a €100 endowment, which they must donate to one of the three candidates at the end of the episode, for a total of €10,000. The candidates are introduced through short, pre-recorded video documentaries—featuring the candidate and a voice-over explaining their predicament—and unstructured studio interviews—during which a panel of two experts briefly questions the candidate about their situation.

ESVV collected and use all 55 episodes ever aired. Together, these provide a diverse set of 165 cases. Normally, a case was represented by one person, and this candidate or their family was the primary beneficiary of the donated money. Occasionally, however, a case was represented by two individuals (20 instances), or a candidate requested money for an outside beneficiary (12 instances). We refer to ESVV for a detailed show description and example cases.

To quantify perceptions of the candidates' situations, ESVV conducted structured rating sessions with 80 Dutch-speaking university students. Each rater evaluated nine cases from three different episodes. For each case, they watched a video clip comprising the candidate's documentary and the subsequent panel interview. Following each video, they provided evaluations along multiple dimensions. First, they assessed the well-being benefit to the candidate of receiving a gift of €100, and that of receiving a gift of €10,000. Using a percentile scale, they reported the candidate's relative rank in terms of overall well-being impact compared to a reference group consisting of the candidate and 99 other individuals randomly drawn from the Dutch population. Next, raters completed the Revised Causal Dimension Scale (CDSII; McAuley, Duncan, and Russell, 1992). This scale includes 12 items, together measuring four attribution dimensions. The personal control dimension measures the extent to which the candidate could control the cause of their situation, and reflects the candidate's responsibility for their financial predicament. The other dimensions are external control (controllability by others), internal locus (cause is internal rather than external to the candidate), and stability (of the cause over time). Before raters provided their CDSII scores, they were asked to briefly describe the cause of the candidate's situation in their own words. Last, they

assessed each candidate's physical appearance on a 1-10 scale, and estimated each candidate's age in years.

For the analysis of their data, ESVV use the Dirichlet-multinomial model, which is well-suited for grouped choice data. The regression results show that the proportion of the prize pool that a candidate received increases with the impact of a donation on well-being, particularly with that of €10,000, and decreases with the candidate's personal control over the cause of their situation, though to a lesser extent. Dominance analysis shows that impact contributes more than twice as much as personal control to explaining donation decisions. The other causal attribution dimensions (external control, internal locus, and stability) and the candidate characteristics age, gender, and physical attractiveness do not significantly affect giving. These results are robust across multiple model specifications and sample restrictions. In sum, using a real-world donation setting with consequential stakes, ESVV's findings suggest that utilitarian considerations outweigh equity-and-desert considerations.

The present paper seeks to replicate and extend these findings with new data, including the donation propensities of experimental subjects and a modified impact measure.

3. Experimental Design

Our replication experiment was conducted as a behavioral pilot study for an fMRI project that aims to use GMG to investigate the neural underpinnings of giving behavior, and served a dual purpose: (i) to examine the robustness of the findings in ESVV, and (ii) to gather supplementary data to guide stimulus selection and experimental design for the neuroimaging study. Appendix B presents screenshots of the experimental instructions and the task interface.

3.1 Stimulus selection

Constraints on total scan time and the need to avoid participant fatigue make it infeasible to include all GMG episodes and cases as stimuli in an fMRI study. We therefore considered a carefully chosen subset, omitting episodes that feature elements that may introduce variability in how subjects interpret and evaluate the donation appeals. Specifically, we excluded episodes in which (i) two individuals appeared together as a pair, representing one case, or (ii) a candidate requested money on behalf of an outside beneficiary, such as a sports, cultural, or societal association.

The resulting set consists of 28 episodes, for a total of 84 distinct cases. Using data from ESVV, Table A1 in Appendix A shows that the key characteristics of this subsample closely resemble those of the full sample. Additionally, Table A2 shows that the regression results for the main model specification from the original study are broadly similar, with personal control appearing somewhat more influential in the restricted sample. In the full sample, impact was estimated to be 2.4 times as important as personal control in explaining the observed donations; in the smaller sample that ratio is 1.8. Possibly, responsibility judgments are more important when candidates request money alone and for themselves rather than as a pair or for an outside beneficiary. In addition, the pseudo R-squared increases from 0.361 for the full sample to 0.544 for the subsample, suggesting that focusing on a more uniform set of cases reduces interpretive and evaluative ambiguity.

3.2 Measures

As in ESVV, subjects watched short video segments from the show and provided ratings on several dimensions. For the purposes of our replication, we focus on the three core constructs: the perceived well-being impact of a donation, the perceived responsibility of the candidate for their situation, and the stated willingness to donate to the candidate.

To measure impact, we asked subjects to rate the extent to which receiving the amount requested by the candidate would improve that candidate's well-being, on a 10-point Likert scale (1 = not at all, 10 = very much). This approach differs from that of ESVV in three ways. First, instead of the Likert scale, ESVV used a comparative percentile scale, asking raters to rank the well-being impact of giving to the candidate relative to that of giving to people in a reference group consisting of the candidate and 99 other individuals randomly drawn from the Dutch population. Second, instead of asking for the impact of the requested amount, they asked for the impact of €100 and for that of €10,000. Last, in the original study, raters were explicitly instructed to consider not only the candidate's wellbeing, but also the well-being of people indirectly involved.

We introduced the changes to reduce complexity and cognitive load. First, subjects are generally more familiar with Likert scales, and using such a scale is inherently less cognitively demanding than a percentile-ranking task. Second, assessing the impact of giving the specific amount requested by the given candidate is easier than assessing that of fixed donation amounts of €100 and €10,000, because it prompts raters to exclusively consider the effect of resolving the full specific problem presented by the candidate, without the need to consider the expected donation decisions of others or the utility of money donated in excess

of the requested amount. From a normative utilitarian perspective, however, it is preferable to compare the effect of a fixed amount across candidates as done by ESVV. ESVV found that €10,000 is more relevant than €100 for measuring well-being impact considerations. Because well-being gains will largely be driven by resolving the candidate's concrete financial shortfall and only minimally by donations exceeding this goal, we expect the perceived well-being impact of the requested amount to be strongly correlated with that of €10,000. Last, the simplification of instructing raters to consider the well-being consequences for the candidate—without any reference to people indirectly affected—likely does not substantially affect their judgments, because in all selected cases the candidate is the primary beneficiary. Descriptive statistics in Subsection 3.4 confirm that there is a strong correlation between this new, simplified impact measure and the two used in ESVV.

Similar to ESVV, we asked raters to assess candidates' responsibility for their situation using the three personal-control dimension items of the Revised Causal Dimension Scale (CDSII; McAuley, Duncan, and Russell, 1992). Each item was rated on a 10-point Likert scale (1 = not at all; 10 = very much). Our implementation differed in two ways from the implementation in the original study. First, we did not ask raters to briefly describe the cause of the candidate's situation in their own words before they provided their ratings. Second, we omitted the other three CDSII dimensions (external control, internal locus, and stability). Both changes were intended to reduce cognitive load and task duration.

To measure raters' willingness to donate, we asked them how likely they would be to donate €100 to the candidate if they were a member of the studio audience. In the instructions, subjects were informed that each audience member had received €100 from the producer and was required to donate this amount in full to one of the candidates, with no option to keep the money for themselves. Responses were given on a continuous slider ranging from 0% ("not at all likely") to 100% ("very likely").

The raters also assessed various other aspects. These included their empathy towards the candidate (both affective and cognitive), their anticipated emotions of giving or not giving (warm-glow and guilt), and their social perceptions of the candidate (competence, warmth, and morality). These additional variables were included to support stimulus selection and experimental design for the fMRI study and are not analyzed here.

3.3 Data collection

The experiment was conducted at the Erasmus Behavioral Lab of Erasmus University Rotterdam, using a PsychoPy-based experimental interface. We recruited 463 Dutch-speaking undergraduate students from

the Rotterdam School of Management (Erasmus University Rotterdam) through the local ERIM Research Participation System that rewards them with course credits. Subjects provided informed consent and were randomly assigned to one of two experimental conditions. After reading condition-specific instructions and completing a brief practice trial, they proceeded to the main task.

Each subject evaluated the cases of five candidates.[2] Subjects were randomly assigned to one of two conditions. In the measure-once condition, they rated each case after watching the full introduction—delivered through a video clip comprising the candidate's documentary and the subsequent panel interview. This matches the setup of ESVV. In the measure-twice condition, subjects first provided ratings after seeing the documentary, and then again after also watching the panel interview. This condition was included to assess the potential for examining belief updating in the fMRI study. The average full video clip was approximately five-and-a-half minutes long (5:23), comprising two-and-a-half minutes of documentary footage (2:31) and three minutes of interview content (2:51).

Here, we collapse the data and exclusively analyze the scores provided after the full introduction of a case. We thus analyze all ratings from the measure-once condition and only the final ratings from the measure-twice condition.[3]

After excluding 22 subjects because of technical issues, the final sample consisted of 441 subjects.[4] Of these, 220 were assigned to the measure-twice condition, and 221 to the measure-once condition. Subjects' average age was 20 years (min: 17, max: 53); 174 (39.5%) identified as female and 267 (60.5%) as male.

3.4 Descriptive statistics

Table 1, Panel A presents summary statistics for the ratings at the rater-case level. With 441 subjects each evaluating five cases, the dataset includes a total of 2,205 ratings per item. On average, the raters assigned an impact score of 7.20 (out of 10), indicating a generally high perceived effectiveness of donating. The average rating for personal control is 4.84 (out of 10), and the average donation propensity is 57.01 (out of 100). Each measure exhibits substantial variability, with standard deviations indicating considerable dispersion and responses that span the full scale.

[2] The assignment of cases was pseudo-randomized to prevent large imbalances in the number of raters per case.

[3] The distributions of the final ratings for the three variables of interest—impact, responsibility, and donation propensity—do not differ significantly between the two conditions (Kolmogorov-Smirnov tests; all $p > 0.215$).

[4] Twelve subjects reported that they could not hear the audio of some stimuli, ten withdrew before completion.

Table 1: Summary statistics

The table presents summary statistics for raters' evaluations of (i) the well-being impact of donating the requested amount, (ii) the candidate's personal control over the cause of their predicament, and (iii) their own propensity to donate €100 if they were an audience member. Each of the 441 raters assessed five different cases from the collection of 84 cases, for a total of 2205 rater-case-level observations and an average of 26.25 raters per case. Panel A reports statistics for the unstandardized scores at the rater-case level. Panel B reports statistics for the case-level averages of the unstandardized scores across all raters who evaluated a given case. Panels C and D report the rater-case-level and case-level statistics after the standardization of (underlying item) scores at the rater level.

	N	Mean	SD	Min	Median	Max
Panel A: Rater-case level, unstandardized						
Impact	2205	7.20	2.03	1.00	7.00	10.00
Personal control	2205	4.84	2.51	1.00	5.00	10.00
Donation propensity	2205	57.01	29.06	0.00	61.00	100.00
Panel B: Case level, unstandardized						
Impact	84	4.84	1.62	2.03	4.74	7.96
Personal control	84	7.21	0.98	4.07	7.29	9.04
Donation propensity	84	57.1	17.09	20.38	59.14	91.35
Panel C: Rater-case level, standardized						
Impact	2175	0.00	0.89	-1.79	0.11	1.79
Personal control	2200	0.00	0.85	-1.79	0.00	1.79
Donation propensity	2205	0.00	0.89	-1.79	0.10	1.78
Panel D: Case level, standardized						
Impact	84	0.00	0.54	-0.96	0.04	0.96
Personal control	84	0.00	0.46	-1.11	0.04	0.79
Donation propensity	84	0.00	0.56	-1.09	0.06	1.12

The average number of raters per case is 26.25, with each case rated by a minimum of 22 and a maximum of 29 raters. To obtain case-level scores, we simply average the scores from all raters who evaluated a given case. Panel B presents the corresponding summary statistics. As expected, averaging reduces the variance, but all three variables nevertheless show substantial variation.

For our analyses, we standardize each rater's score on each item to account for individual differences in how raters interpret the scales. Following ESVV, we compute $\alpha_{ijs}^{ST} = (\alpha_{ijs} - \bar{\alpha}_{is})/\sigma_{is}$, where α_{ijs} denotes rater i's score for candidate j on item s, $\bar{\alpha}_{is}$ is rater i's average score on item s across all candidates, and σ_{is} is the standard deviation of rater i's scores on item s. This standardization procedure yields item scores

that follow a standard normal distribution at the rater level. Panels C and D show summary statistics for the standardized scores at the rater-case level, and for the averages of these scores at the case level.[5,6]

Cronbach's alpha for the standardized scores of three CDSII items that measure personal control is 0.946, indicating high internal consistency. The intraclass correlation coefficients (ICC) for personal control and impact are 0.939 and 0.886, respectively, and suggest high inter-rater agreement.[7] For the measures in ESVV, using the same subset of cases but a different and smaller pool of raters, the three statistics were considerably lower: 0.823 (Cronbach's alpha personal control), 0.694 (ICC personal control) and 0.399 (ICC impact). In the present study, the high Cronbach's alpha value may derive from the absence of the nine items of the other dimensions of the CDSII scale: subjects evaluated the three personal control items in direct succession, with no items from other dimensions in between, and this may have led them to reflect less on each item individually and to assign more similar scores. The higher ICCs are a natural consequence of the larger number of raters per case.

In light of the differences in elicitation, the correlations between the case-level averages of the standardized scores in ESVV and those obtained here are relatively high: our new, simplified impact measure shows a correlation of 0.55 with ESVV's main impact measure (measuring the impact of €10,000; $p < 0.001$), and for the new and original personal control measures, the correlation is 0.74 ($p < 0.001$).[8] The weaker correlation for impact than for personal control is likely primarily attributable to the relatively low ICC of the impact measure employed in ESVV.

[5] In Panel C, the average score is zero for each of the variables by construction. Note that the standard deviations are below one; standardizing within the small rater-level subsets effectively eliminates extremes, and this translates into the underrepresentation of tail observations at the aggregate level relative to the standard normal distribution.

[6] Six raters assigned the same impact score to each of their five cases, and one did not vary their scores for the three personal control items across their five cases. Standardization is impossible is impossible in the absence of variation, and these occurrences therefore result in a lower number of observations for the standardized rater-case-level impact and personal control variables.

[7] Following ESVV, we computed $\mathrm{ICC}(2,k) = \sigma^2_{case}/(\sigma^2_{case} + (\sigma^2_{residual}/k))$. We used a two-way random-effects model that decomposes the overall variance into between-case variance and residual variance, and set k at the average number of raters per case.

[8] The correlation between the present impact measure and ESVV's impact of €100 measure is 0.35 ($p = 0.001$).

4. Analyses and results

4.1 Donations in the show

We first analyze the actual giving behavior on the show, using the new ratings of impact and personal control. The next subsection considers the stated donation propensities of the individual raters themselves.

Figure 1 shows ternary plots for the percentage of the total prize pool of €10,000 the audience allocated to the candidates who rank highest, middle, or lowest in terms of raters' perceptions of impact (Panel A) and personal control (B). The points are tilted towards the top and left, and to the right and bottom, respectively, indicating that more money went to those who would benefit more from having their problem resolved, and to those with less personal control over its cause. The average percentages displayed next to the vertices show that the differences between the lowest- and highest-ranked candidates are substantial: on average, the candidate ranked highest in terms of impact (personal control) received 25 (28) percentage points more (less) than the candidate ranked lowest. The results from this first, descriptive analysis align with those in ESVV.

Following ESVV, we use multivariate Dirichlet-multinomial regression to disentangle the effects of impact and personal control and to control for other variables. The Dirichlet-multinomial model accounts for unobserved correlations in audience members' choices, and ensures that donation probabilities sum to one and reflect candidates' characteristics relative to others in the same episode.

Table 2, Panel A presents regression results for four sets of explanatory variables. Model 1 exclusively considers the new impact and personal control measures. Models 2-4 add variables from ESVV: Model 2 adds the demographic variables for age and gender, Model 3 also adds the attractiveness ratings, and Model 4 further expands and completes the model with the CDSII dimensions external control, internal locus, and stability. To facilitate interpretation and comparison, all continuous explanatory variables are cross-sectionally standardized.

Depending on the set of control variables, a one-standard-deviation increase in the new impact measure increases a candidate's share of the prize pool by a statistically significant 6.9 to 9.5 percentage points (all $p < 0.001$). For the new personal control measure, the estimated effect is negative and ranges from -4.5 ($p = 0.006$) to -1.0 percentage points ($p = 0.702$).

As in ESVV, we use general dominance analysis to evaluate the relative contributions of impact and personal control to the log-likelihood. Table 2, Panel B shows that the contribution of impact is between 1.38 (Model 1) and 1.95 (Model 4) times that of personal control.

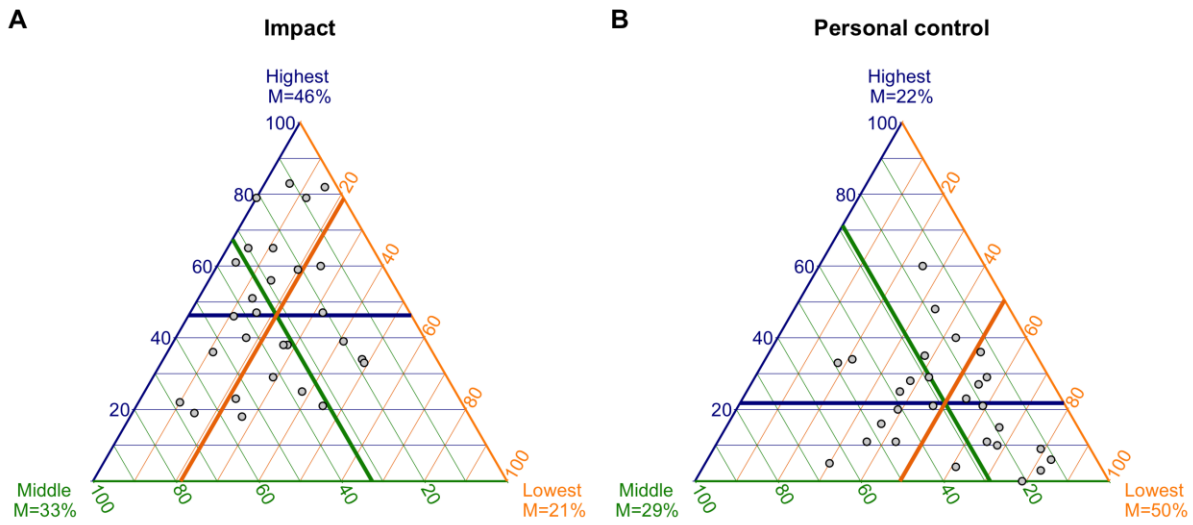

Figure 1: Ternary plots of donations in the show by rankings on impact and personal control. The figure shows ternary plots for the percentage shares received by the candidates in the show ranking highest, middle, or lowest on *Impact* (Panel A) and *Personal control* (B) according to the case-level average ratings obtained from the replication experiment. Both variables are after the standardization of (underlying item) scores at the rater level.

The right-hand part of Table 2 shows the results that we obtain for the exact same models when all variables—also the impact and personal control ratings—are taken from ESVV, allowing for a clean comparison of how well the results replicate the original findings for the subset of 28 episodes that we consider. Collectively, the two new measures improve the empirical fit of the models by a substantial 10-13 percentage points. This improvement likely derives from the greater number of raters per case, which increased the reliability of the measures. With the new measures, the estimated coefficients for impact are slightly more positive (across all models) and those for personal control are somewhat less negative (Models 2-4), while the ratios of the contributions of impact and personal control are roughly similar (all models). The attenuation of the personal control coefficient may result from changes in the relationships between the different variables: it is possible that subjects evaluated the personal control dimension somewhat differently in the absence of exposure to the items for the other CDSII dimensions, factoring in aspects they might otherwise have associated with the omitted dimensions. Replacing the impact and personal control variables barely affects the coefficients of any of the six control variables and their contributions to the empirical fit.

Table 2: Dirichlet-multinomial regression results for donations in the show

The table presents the average marginal effects from Dirichlet-multinomial regression analyses (Panel A), and the outcomes of the corresponding general dominance analyses (Panel B). The dependent variable is the proportion of the prize pool that a candidate received. For the first set of four models, the variables *Impact* and *Personal control* are obtained from the replication experiment; for the second set, these variables are from ESVV. The other variables in both sets are from ESVV. With the exception of *Age* and *Female*, all explanatory variables are after the standardization of (underlying item) scores at the rater level. Additionally, all continuous explanatory variables are cross-sectionally standardized. *Pseudo R^2* is the Efron pseudo R-squared, which is the sum of the squared model residuals divided by the total variability in the dependent variable. Standard errors are in parentheses. Asterisks denote statistical significance at the one ***, five **, and ten * percent level, respectively.

	New measures				Original measures			
	Model 1	Model 2	Model 3	Model 4	Model 1'	Model 2'	Model 3'	Model 4'
Panel A: Regression results								
Impact	0.069***	0.080***	0.086***	0.095***	0.066***	0.074***	0.080***	0.081***
	(0.016)	(0.018)	(0.017)	(0.019)	(0.014)	(0.015)	(0.015)	(0.015)
Personal control	-0.045***	-0.033*	-0.028	-0.010	-0.044***	-0.042***	-0.041***	-0.055***
	(0.016)	(0.018)	(0.017)	(0.025)	(0.013)	(0.013)	(0.012)	(0.018)
External control				-0.004				-0.002
				(0.016)				(0.016)
Internal locus				-0.025				-0.020
				(0.017)				(0.016)
Stability				0.007				-0.022
				(0.015)				(0.018)
Age		-0.015	-0.004	-0.012		-0.021	-0.009	-0.014
		(0.014)	(0.014)	(0.015)		(0.016)	(0.017)	(0.017)
Female		0.018	-0.001	-0.005		-0.009	-0.034	-0.046
		(0.027)	(0.028)	(0.027)		(0.032)	(0.034)	(0.033)
Attractiveness			0.029**	0.022			0.028*	0.024
			(0.013)	(0.013)			(0.016)	(0.016)
Log-likelihood	-210.1	-209.2	-206.9	-205.5	-217.2	-216.3	-214.8	-212.9
Pseudo R^2	0.571	0.577	0.609	0.642	0.438	0.456	0.481	0.544
No. cases	84	84	84	84	84	84	84	84
No. episodes	28	28	28	28	28	28	28	28
Panel B: Dominance analysis results								
Impact	57.9	57.1	56.8	53.1	59.4	55.7	54.6	48.8
Personal control	42.1	36.2	32.2	27.3	40.6	36.4	33.4	27.1
External control				1.0				1.0
Internal locus				5.3				5.8
Stability				4.5				7.0
Age		2.2	1.2	1.4		2.6	1.2	1.3
Female		4.4	3.4	2.7		5.4	5.1	4.6
Attractiveness			6.4	4.7			5.8	4.4
Impact / Personal control	1.38	1.58	1.76	1.95	1.46	1.53	1.63	1.80

The conclusions from the analysis of the audience donations with the new measures are that giving increases more with impact than it decreases with personal control, and that impact contributes considerably more to explaining donation decisions than personal control. These conclusions are identical to the conclusions in ESVV.

4.2 Donation propensities raters

The benefits of the analysis of the giving decisions of the audience, as presented in the previous subsection, are that the decisions are real and consequential for the candidates. A limitation, however, is that the donations are observed only at the aggregate level. Also, audience members' perceptions of impact and personal control are unknown, and the evaluations from the independent raters may not fully align with these.

In the present subsection we turn to a more granular analysis, and consider the stated willingness to donate of the raters in the replication experiment. We use raters' assessments of the well-being impact of donating and the candidate's control over the cause of their predicament, and examine how these perceptions account for differences in their own donation propensities.

Raters' willingness to donate to a given candidate closely mirrors the actual donations of the audience on the show. Regressing the proportion of the prize pool that a candidate received on raters' average standardized stated donation propensity—using the Dirichlet-multinomial model—results in an Efron pseudo R-squared of 0.652. This single variable thus provides greater explanatory power than the best-fitting multivariate model in Table 2.

Figure 2, Panels A and B present scatterplots of raters' individual donation propensity as a function of their perceptions of the impact of donating on well-being (A) and the candidate's control over the cause of their situation (B), along with the coefficient and R-squared value of the fitted linear regression lines. Panels C and D display bar charts of the average donation propensity across deciles of perceived impact (C) and personal control (D). Clearly, rater's stated willingness to donate increases with impact and decreases with personal control. Consistent with the results for the audience donations in the show, these descriptive univariate analyses suggest that giving increases more with impact than it decreases with personal control, and that the explanatory power of impact is greater than that of personal control.

The multivariate analysis of audience members' choices required the use of the Dirichlet-multinomial model. Here, we can use simple OLS regression to disentangle the different factors. Unlike the grouped

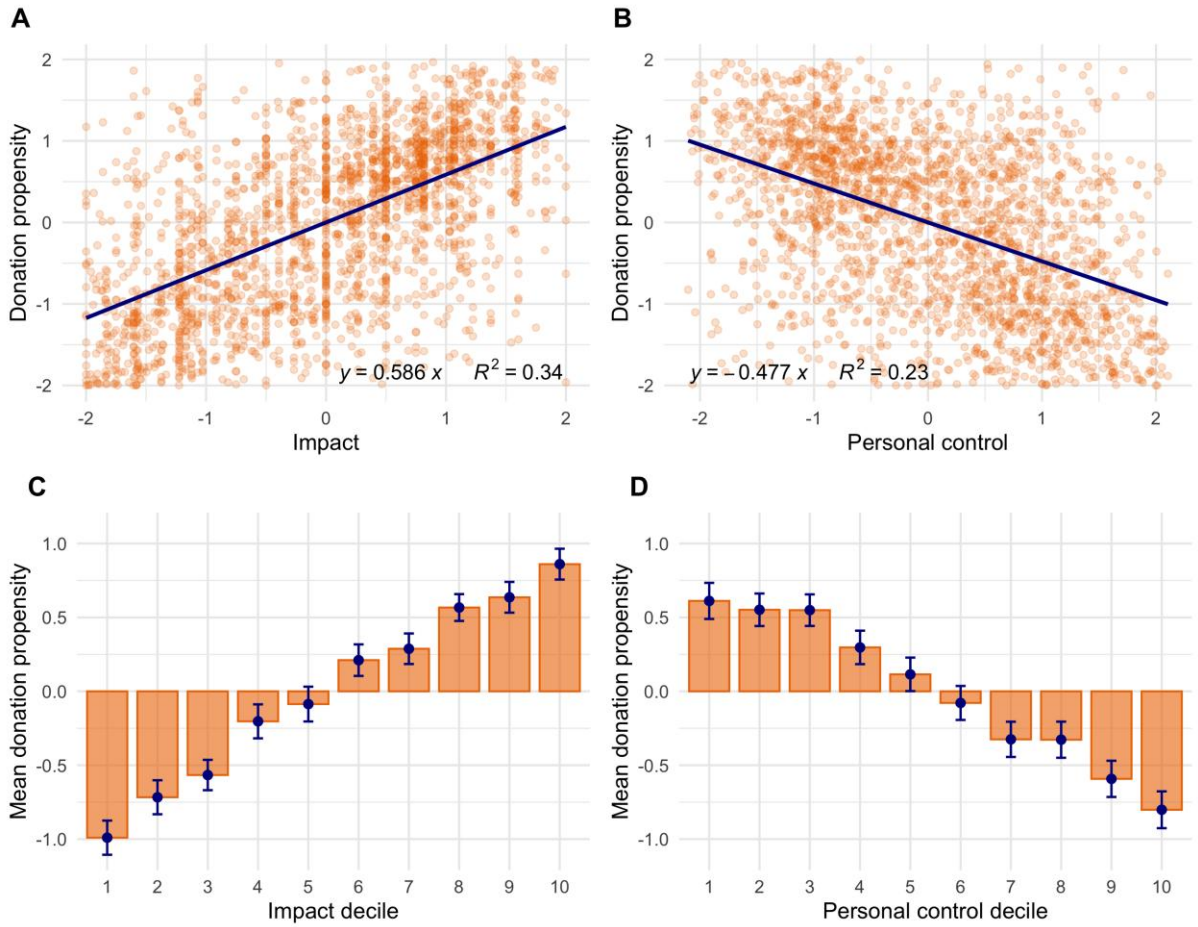

Figure 2: Raters' donation propensities by impact and personal control. Panels A and B present scatterplots of raters' individual donation propensity as a function of their perceived impact (A) and perceived personal control (B), along with the fitted linear regression lines. Panels C and D display bar charts of raters' average donation propensity across deciles of perceived impact (C) and perceived personal control (D), with error bars indicating 95% confidence intervals. All variables are after the standardization of (underlying item) scores at the rater level, and then cross-sectionally standardized to ensure a mean of zero and a standard deviation of one.

choice data for the audience, the donation propensities of the raters in the replication experiment are measured at the individual level. Also, raters evaluated five pseudo-randomly selected cases, eliminating interdependency between cases from the same episode.

Table 3, Panel A presents the regression results. Model 1 only includes raters' impact and personal control assessments, and Models 2-4 add case-specific control variables from ESVV. All continuous variables are cross-sectionally standardized, and standard errors are clustered at both the rater and case level.

Table 3: Regression results for raters' donation propensities

The table reports the OLS regression results (Panel A), and the outcomes of the corresponding general dominance analyses (Panel B). The dependent variable is the stated donation propensity of a rater towards a particular case. *Impact* and *Personal control* are from the replication experiment and at the rater-case level; all other variables are from ESVV and at the case level. The first set of five models is estimated for all raters combined, the remaining models are estimated for female (4F and 5F) or male (4M and 5M) raters only. With the exception of *Age* and *Female*, all variables are after the standardization of (underlying item) scores at the rater level. Additionally, all continuous variables are cross-sectionally standardized. Standard errors are clustered at both the rater and case level, and shown in parentheses. Asterisks denote statistical significance at the one ***, five **, and ten * percent level, respectively.

	All raters					Female (F) or Male (M) raters only			
	Model 1	Model 2	Model 3	Model 4	Model 5	Model 4F	Model 4M	Model 5F	Model 5M
Panel A: Regression results									
Impact	0.482***	0.473***	0.472***	0.468***	0.386***	0.491***	0.452***	0.405***	0.377***
	(0.026)	(0.025)	(0.024)	(0.023)	(0.023)	(0.037)	(0.031)	(0.033)	(0.030)
Personal control	-0.319***	-0.315***	-0.315***	-0.289***	-0.240***	-0.271***	-0.299***	-0.247***	-0.236***
	(0.028)	(0.028)	(0.028)	(0.027)	(0.024)	(0.040)	(0.032)	(0.037)	(0.031)
External control				-0.006		-0.024	0.005		
				(0.044)		(0.054)	(0.050)		
Internal locus				-0.155***		-0.130**	-0.170***		
				(0.044)		(0.052)	(0.047)		
Stability				0.048		0.056	0.046		
				(0.038)		(0.048)	(0.040)		
Age		0.068*	0.064*	0.022		0.025	0.022		
		(0.037)	(0.037)	(0.034)		(0.039)	(0.040)		
Female		0.099	0.105	0.041		0.113	-0.002		
		(0.068)	(0.070)	(0.064)		(0.072)	(0.076)		
Attractiveness			-0.012	-0.049		-0.046	-0.050		
			(0.043)	(0.039)		(0.053)	(0.038)		
Constant	0.000	-0.068	-0.072	-0.028	0.000	-0.081*	0.004	0.000	0.000
	(0.031)	(0.049)	(0.054)	(0.050)	(0.007)	(0.047)	(0.061)	(0.012)	(0.010)
Case fixed effects	NO	NO	NO	NO	YES	NO	NO	YES	YES
R^2	0.435	0.440	0.440	0.459	0.548	0.464	0.459	0.582	0.560
Observations	2,170	2,170	2,170	2,170	2,170	860	1,310	860	1,310
No. raters	434	434	434	434	434	172	262	172	262
No. cases	84	84	84	84	84	84	84	84	84

Panel B: Dominance analysis results									
Impact	63.4	61.0	60.9	56.2	57.0	58.7	54.0	56.5	55.9
Personal control	36.6	35.1	35.1	30.1	35.5	28.4	31.0	34.6	35.9
External control				0.8		0.6	1.0		
Internal locus				6.1		3.6	8.0		
Stability				3.9		4.3	3.7		
Age		1.7	1.5	0.9		0.6	1.1		
Female		2.2	2.3	1.6		3.2	0.9		
Attractiveness			0.2	0.4		0.5	0.3		
Case fixed effects					7.5			8.9	8.2
Impact / Personal control	1.73	1.74	1.73	1.87	1.61	2.07	1.74	1.63	1.56

Depending on the set of control variables, a one-standard-deviation increase in a rater's perceived impact increases their donation propensity by a statistically significant 0.47 or 0.48 standard deviations (all $p < 0.001$). For personal control, the effect is statistically significantly negative throughout, and estimated at -0.32 or -0.29 standard deviations (all $p < 0.001$).

We again use general dominance analysis to evaluate the relative contributions of impact and personal control to the empirical fit of the different models. Because of the switch to OLS regression, we now consider the variables' contributions to the R-squared (rather than to the log-likelihood). Panel B shows that the contribution of impact is between 1.73 (Model 1 and 3) and 1.87 (Model 4) times that of personal control.

Model 5 exclusively considers how between-rater variations in donation propensity vary with their impact and personal control perceptions, using fixed effects to eliminate systematic differences in donation propensities across cases. Despite some slight attenuation, impact and personal control remain strong predictors, and impact remains the more important of the two. The results indicate that, for a given case, a rater who perceives a higher impact of donating reports a higher donation propensity. The reverse holds for personal control: the more strongly a rater believes that the given candidate had control over the cause of their predicament, the less they are inclined to give to that person. According to the dominance analysis, the contribution of impact to the model's empirical fit is 1.61 times that of personal control.

The individual-level donation propensities also allow for examining whether there are gender differences in how men and women weigh impact and personal control. Models 4F and 4M show the results for female and male raters separately, when all case-specific control variables from ESVV are included. For both impact and personal control, the estimated coefficients do not differ significantly between male and female raters (*Impact*: $p = 0.450$; *Personal control*: $p = 0.593$). Models 5F and 5M similarly separate the raters by gender, but instead use case fixed effects. Again, the coefficients for male and female raters are similar and statistically not significantly different (*Impact*: $p = 0.523$; *Personal control*: $p = 0.831$).

Altogether, the results for raters' stated individual donation propensities closely mirror those for the giving behavior of the audience on the show, and further validate the robustness of the findings reported in ESVV.

5. Conclusion

This replication and extension study reaffirms the key findings of ESVV. As in ESVV, we use the Dutch TV show *Geld Maakt Gelukkig*, in which an audience of one hundred allocates a real €10,000 among three

candidates who face a problem beyond their financial capacity. For our replication, we collected new impact and personal control ratings, and, as an extension, we also elicited raters' own donation propensities.

Using the new ratings, the first part of our analysis replicates ESVV's finding that giving in the show increases more with perceived impact than it decreases with perceived personal control, and that impact accounts for a greater share of explanatory power in the audience's donation decisions. In the second analysis, we link raters' own donation propensities to their perceptions of impact and personal control. A key distinction from the audience's donations—which occurred in a joint-evaluation context where they chose between three candidates—is that raters' propensities were elicited in a separate-evaluation context, where each case was judged in isolation. Again, we find that impact is the stronger predictor. Additionally, we find no evidence that male and female raters' donation propensities are differentially influenced by impact and personal control, despite previous research suggesting gender differences in moral views (Atari, Lai, and Dehghani, 2020).

We introduced several methodological changes relative to ESVV. First, we simplified the measure of impact to reduce raters' cognitive load. Second, we increased the number of raters per case, yielding higher inter-rater reliability for the two variables. Third, we changed the structure of the evaluation task: unlike ESVV, where raters viewed cases in the same order as the audience, our raters assessed cases in random order. Despite these changes, we observe the same qualitative and quantitative patterns.

The confirmation of the results in ESVV is important, as these are somewhat surprising in the light of the picture emerging from the literature. Even though there exist no other direct empirical comparisons, prior studies generally show strong empirical support for the importance of equity and desert considerations, whereas the evidence for the importance of utilitarianism is mixed. In the social sciences, there is increasing awareness that surprising results require compelling evidence (Ioannidis, 2005; Maniadis, Tufano, and List, 2014).

Taken together, the analyses in ESVV and in the present paper confirm that both utilitarian and equity-and-desert-based concerns shape giving behavior, and show that utilitarian impact considerations tend to dominate. The findings contribute to the understanding of a fundamental aspect of human decision-making and social interactions, and can inform the development of policies aimed at enhancing economic and social well-being. Future research may, for example, compare the roles of impact and responsibility considerations in settings different than ours, and for culturally different populations.

References

Atari, Mohammad, Mark H. C. Lai, and Morteza Dehghani. 2020. “Sex differences in moral judgements across 67 countries.” *Proceedings of the Royal Society B: Biological Sciences* 287(1937), 20201201.

Eberhardt Hiabu, Inka, Paul Smeets, Martijn J. van den Assem, and Dennie van Dolder. 2026. “Impact or responsibility? Giving behaviour in a televised natural experiment.” *Economic Journal*, forthcoming, available at https://doi.org/10.1093/ej/ueag038.

Ioannidis, John P. A. 2005. “Why most published research findings are false.” *PLoS Medicine* 2(8): 696-701.

Maniadis, Zacharias, Fabio Tufano, and John A. List. 2014. “One swallow doesn’t make a summer: New evidence on anchoring effects.” *American Economic Review* 104(1): 277–290.

McAuley, Edward, Terry E. Duncan, and Daniel W. Russell. 1992. “Measuring causal attributions: The revised causal dimension scale (CDSII).” *Personality and Social Psychology Bulletin* 18(5): 566-573.

Appendix A: Comparisons full set and subset of cases

Table A1: Summary statistics for the full set and for the subset of cases

The table displays summary statistics for variables describing the candidates and their cases for the full set of 165 cases (55 episodes) and for the subset of 84 cases (28 episodes) employed in the present paper. All data are from ESVV, and definitions are as in Table 2 in ESVV.

	Full set		Subset	
	Mean	SD	Mean	SD
Candidate characteristics				
Age	38.16	13.27	38.97	12.99
Female	0.66	0.47	0.69	0.47
White	0.94	0.24	0.95	0.21
Attractiveness	5.14	1.35	4.95	1.32
Money amounts				
Amount requested	5,087	2,295	4,845	2,348
Amount received	3,333	1,767	3,333	1,968
Impact				
Impact €100	38.07	13.78	40.62	13.27
Impact €10,000	67.57	14.11	68.00	14.19
Causal attributions				
Personal control	4.80	1.48	4.67	1.42
External control	4.98	1.35	5.00	1.36
Internal locus	5.19	1.60	5.17	1.68
Stability	4.65	1.41	4.71	1.30
No. cases	165		84	
No. episodes	55		28	

Table A2: Dirichlet-multinomial regression results for the full set and for the subset of cases

The table presents average marginal effects from Dirichlet-multinomial regression analyses (Panel A) and the outcomes of the corresponding general dominance analyses (Panel B), both for the full set of 165 cases (55 episodes) and for the subset of 84 cases (28 episodes) employed in the present paper. The regression specification corresponds to Model 6 in ESVV, which is their main specification. All data are from ESVV, and definitions are as in Table 3 in ESVV. Standard errors are in parentheses. Asterisks denote statistical significance at the one ***, five **, and ten * percent level, respectively.

	Full set	Subset
Panel A: Regression results		
Impact €10,000	0.061***	0.081***
	(0.011)	(0.015)
Personal control	-0.027**	-0.055***
	(0.013)	(0.018)
External control	-0.018	-0.002
	(0.012)	(0.016)
Internal locus	-0.009	-0.020
	(0.012)	(0.016)
Stability	0.001	-0.022
	(0.014)	(0.018)
Age	-0.011	-0.014
	(0.012)	(0.017)
Female	-0.012	-0.046
	(0.023)	(0.033)
Attractiveness	-0.002	0.024
	(0.011)	(0.016)
Log-likelihood	-428.4	-212.9
Pseudo R^2	0.361	0.544
No. cases	165	84
No. episodes	55	28
Panel B: Dominance analysis results		
Impact €10,000	57.6	48.8
Personal control	23.7	27.1
External control	2.7	1.0
Internal locus	3.6	5.8
Stability	7.4	7.0
Age	1.0	1.3
Female	3.6	4.6
Attractiveness	0.4	4.4
Impact / Personal control	2.43	1.80

Appendix B: Screenshots Experiment

The figures below present screenshots from the experiment. Apart from the consent screens, all text was in Dutch. English translations are available upon request.

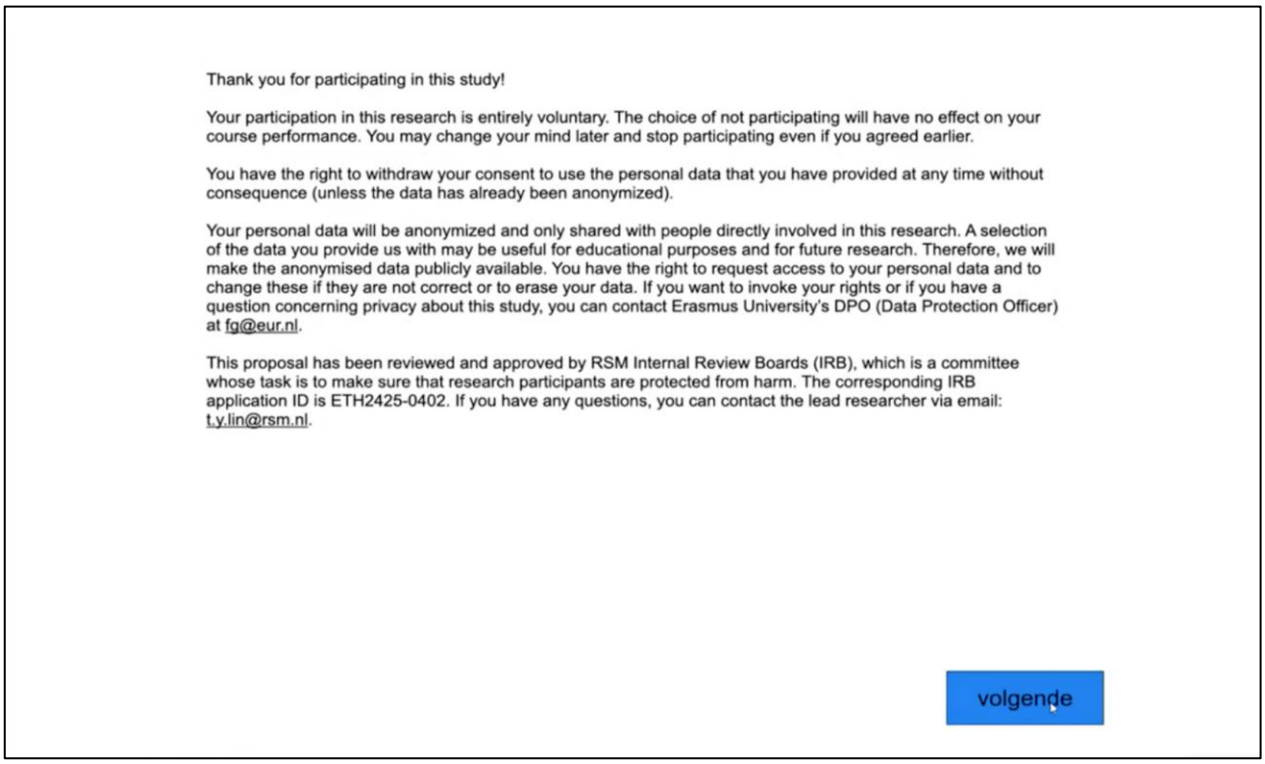

Figure B1: Screenshot consent (1/2)

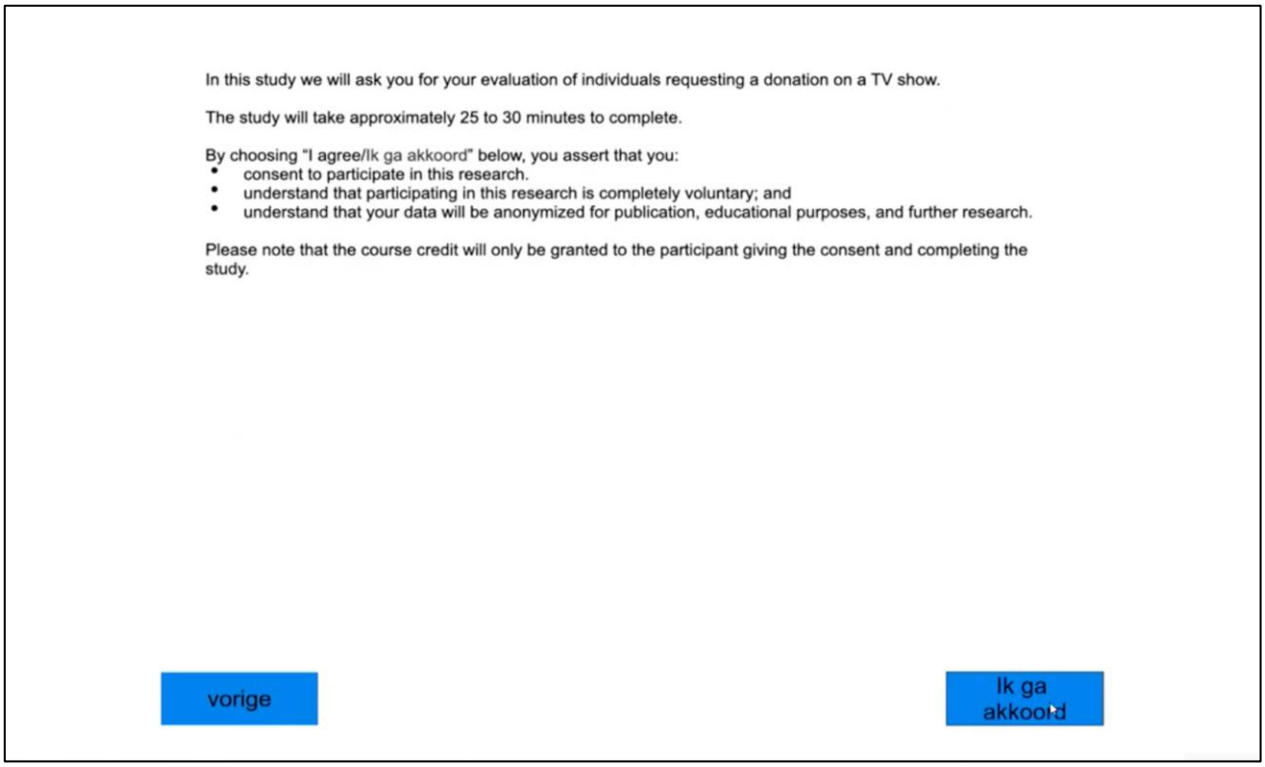

Figure B2: Screenshot consent (2/2)

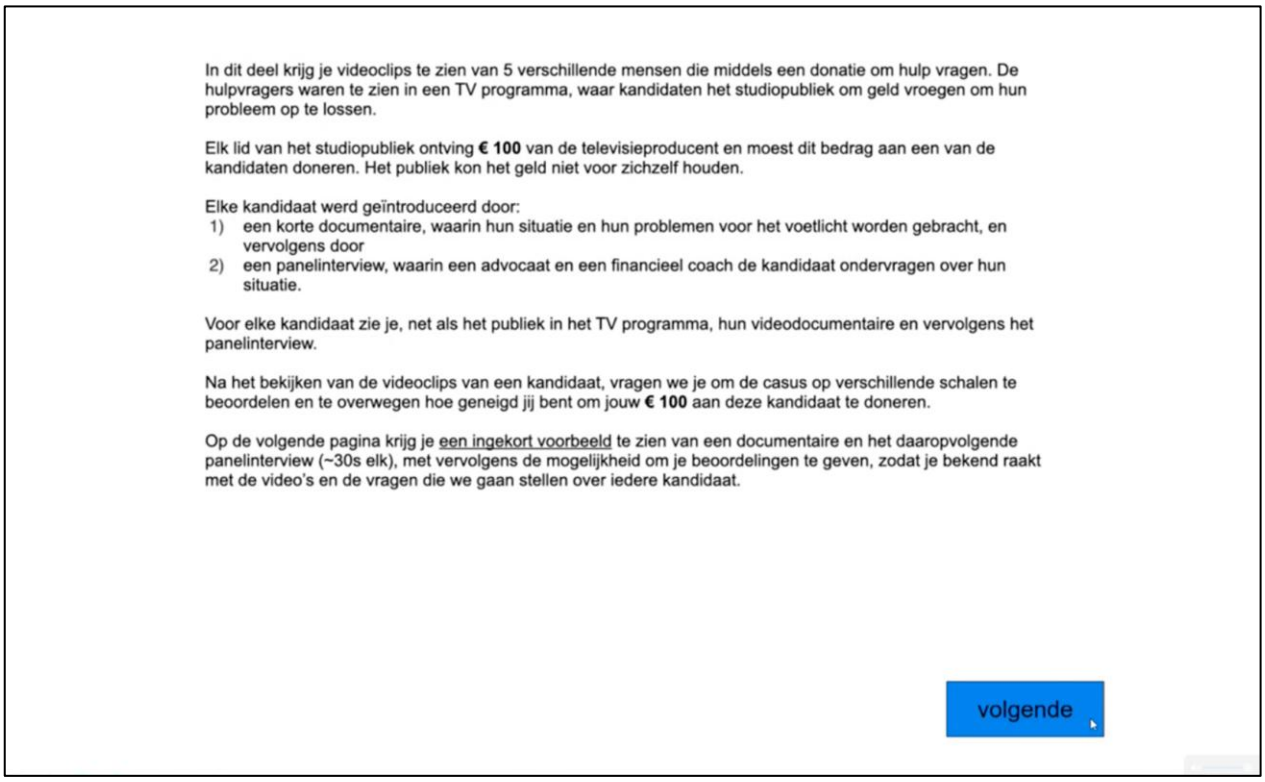

Figure B3A: Screenshot general instructions (measure-once condition only)

In dit deel krijg je videoclips te zien van 5 verschillende mensen die middels een donatie om hulp vragen. De hulpvragers waren te zien in een TV programma, waar kandidaten het studiopubliek om geld vroegen om hun probleem op te lossen.

Elk lid van het studiopubliek ontving **€ 100** van de televisieproducent en moest dit bedrag aan een van de kandidaten doneren. Het publiek kon het geld niet voor zichzelf houden.

Elke kandidaat werd geïntroduceerd door:
1) een korte documentaire, waarin hun situatie en hun problemen voor het voetlicht worden gebracht, en vervolgens door
2) een panelinterview, waarin een advocaat en een financieel coach de kandidaat ondervragen over hun situatie.

Voor elke kandidaat zie je, net als het publiek in het TV programma, hun videodocumentaire en vervolgens het panelinterview.

Na het bekijken van de videoclips van een kandidaat, vragen we je om de casus op verschillende schalen te beoordelen en te overwegen hoe geneigd jij bent om jouw **€ 100** aan deze kandidaat te doneren. Dit doen we op twee momenten: zowel na de documentaire als na het panelinterview.

Voor de metingen na de **documentaire**, vragen we je om dit te doen **op basis van de inhoud van de documentaire**.

Voor de metingen na het **panelinterview**, vragen we je om dit te doen **op basis van zowel de documentaire als het interview**.

Op de volgende pagina krijg je een ingekort voorbeeld te zien van een documentaire en het daaropvolgende panelinterview (~30s elk), met vervolgens de mogelijkheid om je beoordelingen te geven, zodat je bekend raakt met de video's en de vragen die we gaan stellen over iedere kandidaat.

volgende

Figure B3B: Screenshot general instructions (measure-twice condition only)

Figure B4: Screenshot video clip

Figure B5: Screenshot elicitation (1/6)

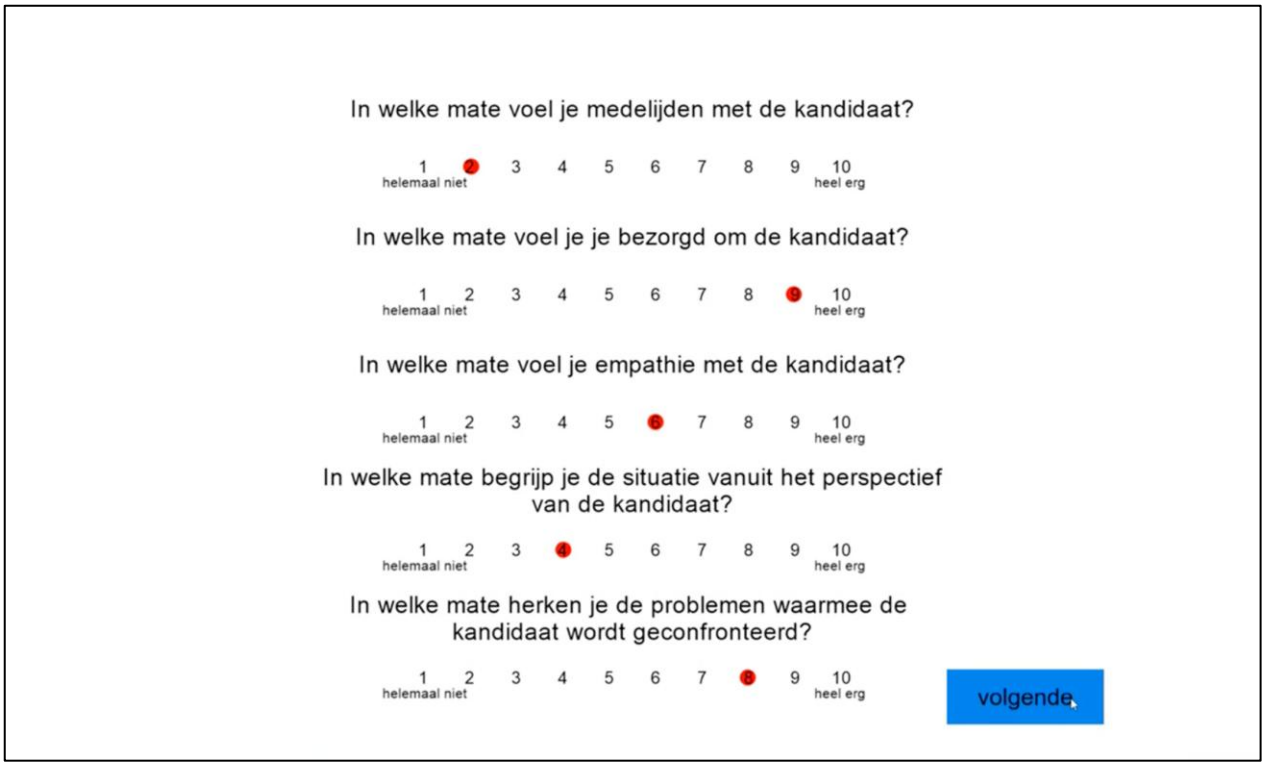

Figure B6: Screenshot elicitation (2/6)

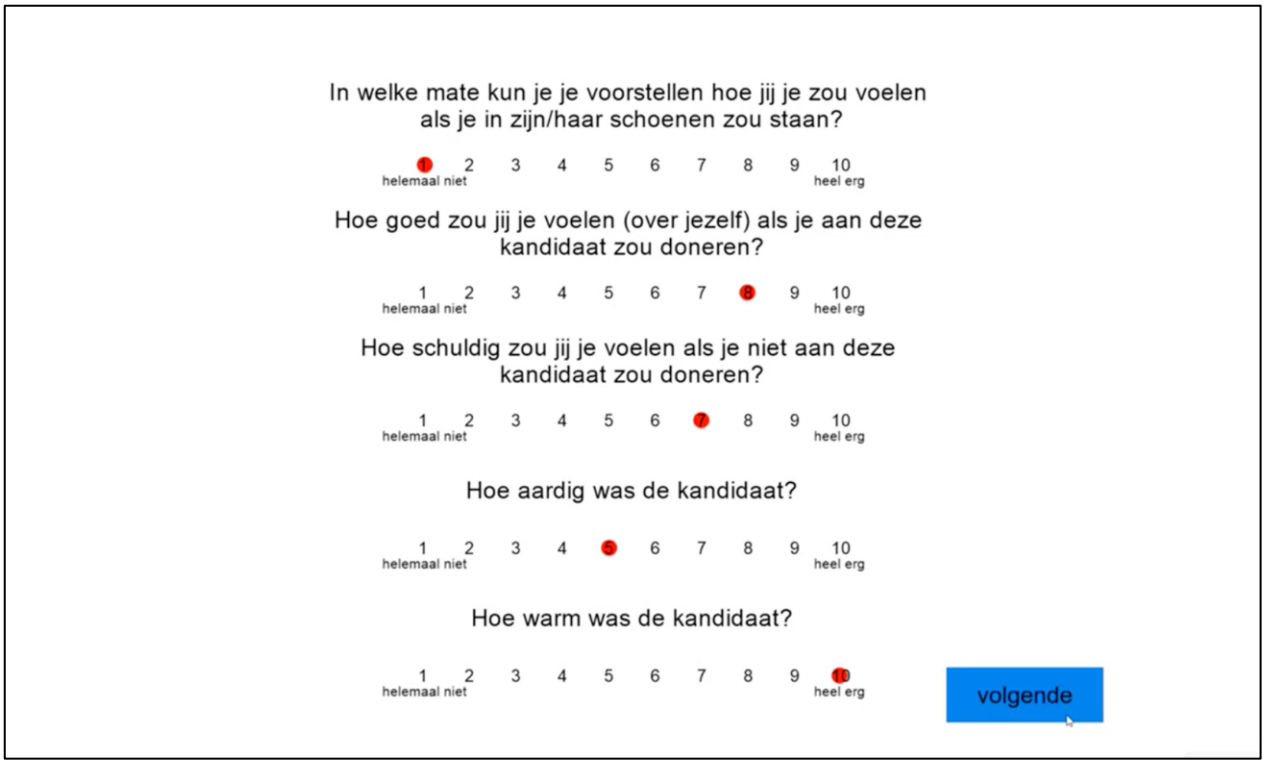

Figure B7: Screenshot elicitation (3/6)

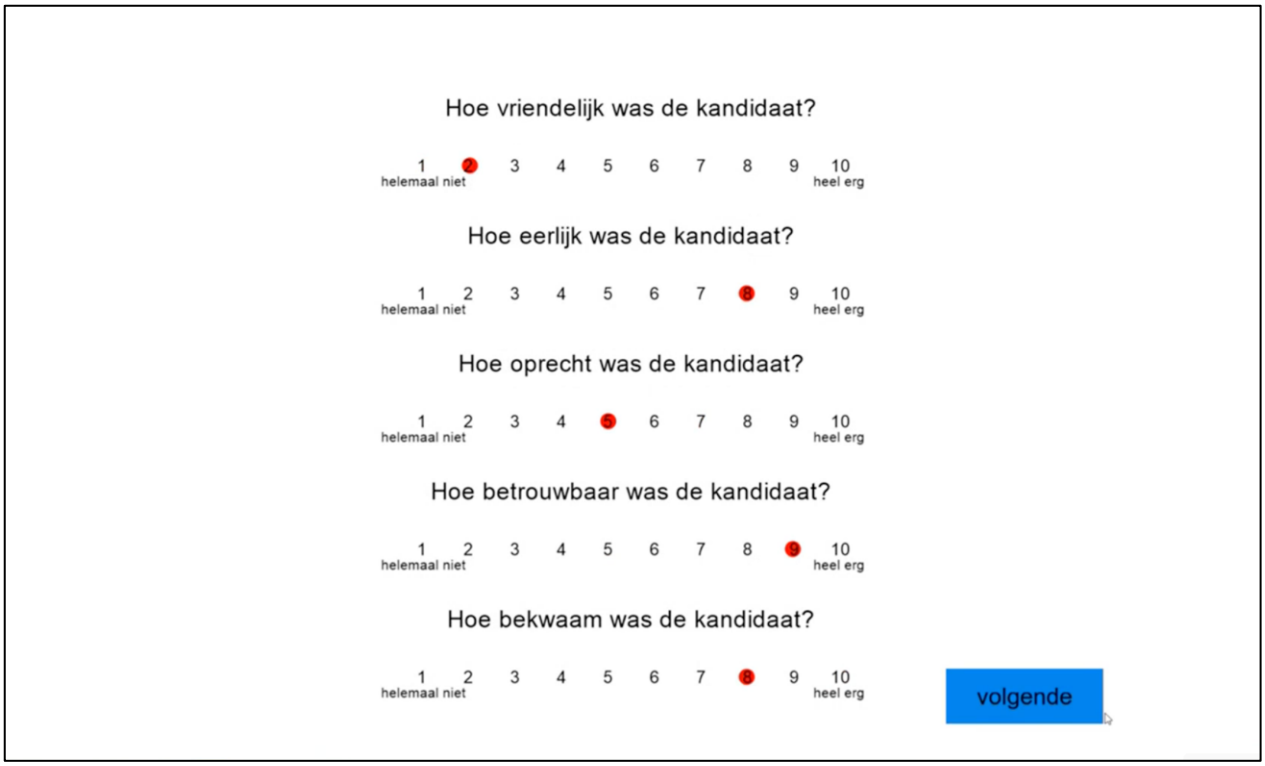

Figure B8: Screenshot elicitation (4/6)

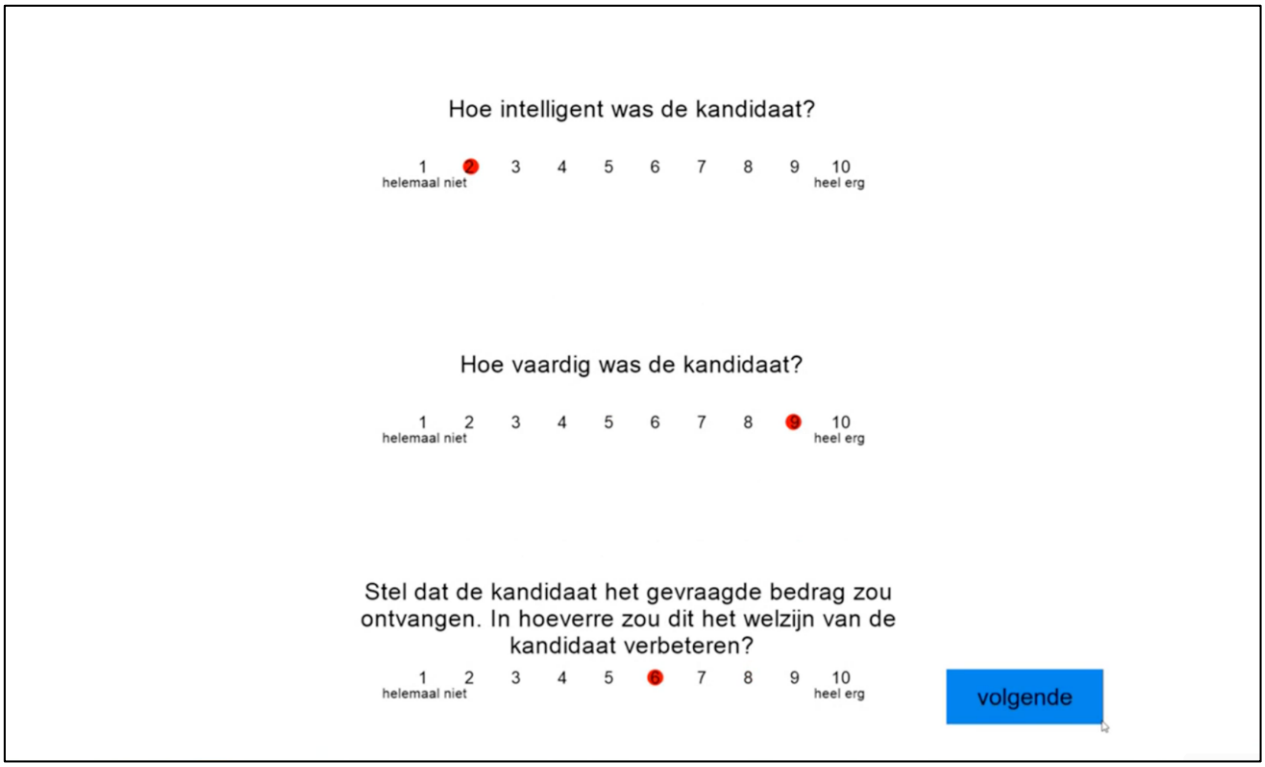

Figure B9: Screenshot elicitation (5/6)

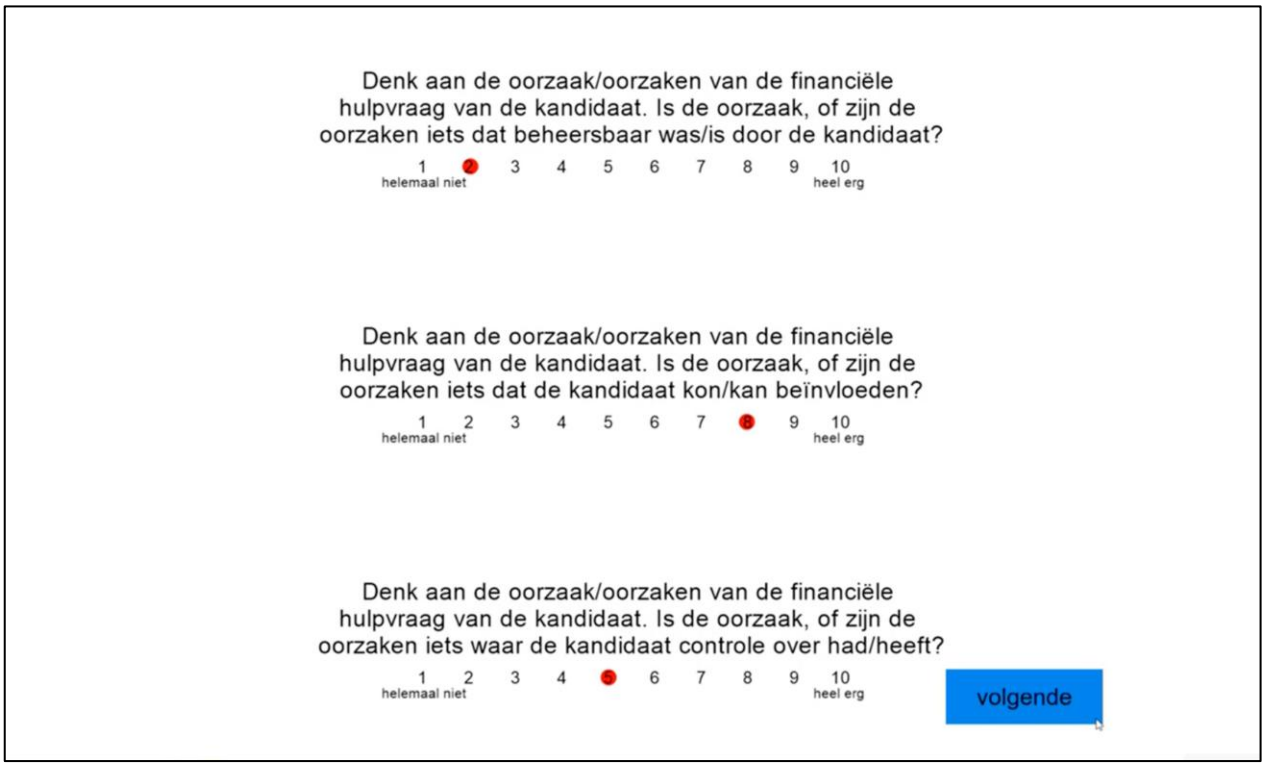

Figure B10: Screenshot elicitation (6/6)

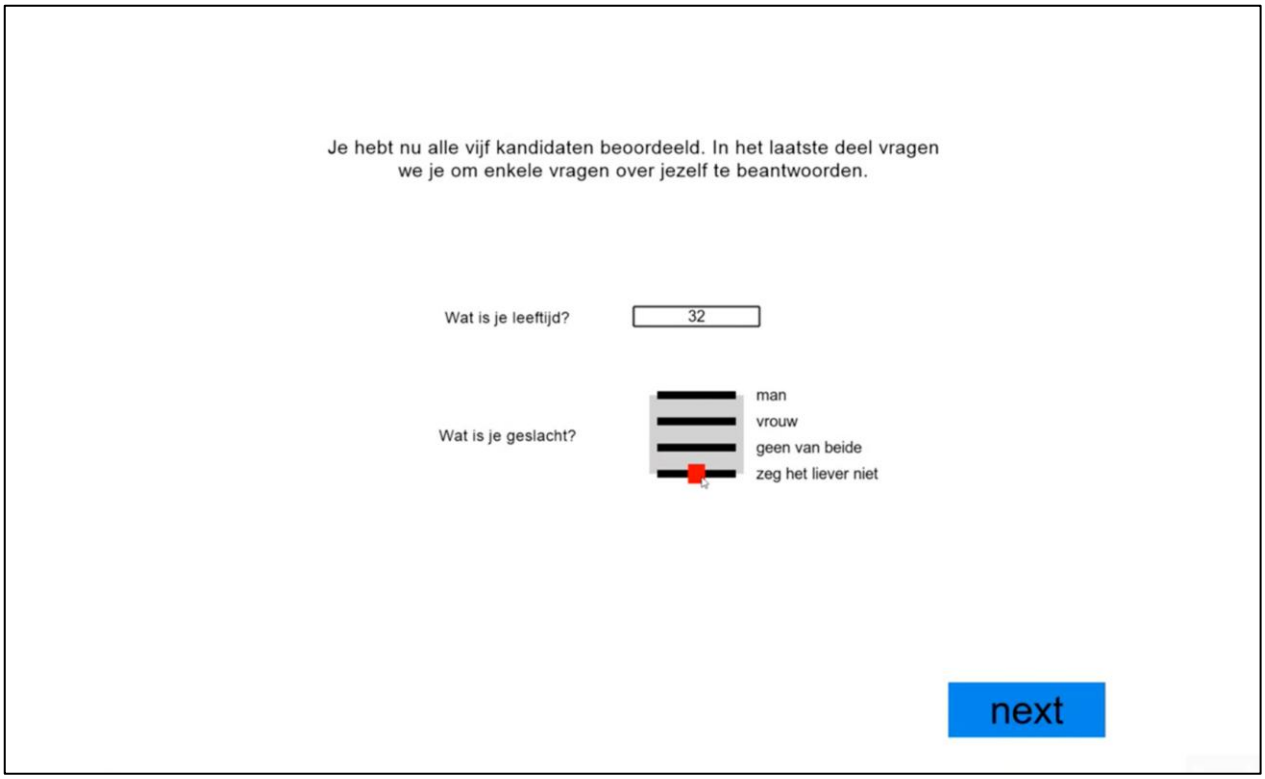

Figure B11: Screenshot demographics

Figure B12: Screenshot end